AALBORG UNIVERSITY
DENMARK

How can internationally operating SMEs successfully integrate Big Data with solutions of external providers?

Philipp Maronitsch

Supervisor: Mohammad Bakhtiar Rana

Master Thesis

2021

AALBORG UNIVERSITY
DENMARK

STANDARDTITLEPAGEFORPROJECTS

To be completed by the student(s)

Subjects	Project:	Thesis: X	Written Assignment:

Study programme:	International Business	
Semester:	4	
Exam Title:	Master Thesis	
Group Number:	-	
Names + Student Nos of group member(s):	Name(s)	Student Number(s)
	Philipp Maronitsch	20192042
Submission date:	June 2nd, 2021	
Project Title /Thesis Title	How can internationally operating SMEs successfully integrate Big Data with solutions of external providers?	
According to module description, maximum number of characters/words/pages of the paper	60 pages à 2,400 characters (+/- 5% tolerance)	
Number of characters/words/pages *(Standard page = 2400 characters including Tables and Figures, excluding References, Appendices, Front Page, Table of Contents)*	149.745 characters / 62 pages	
Supervisor (project/thesis):	Mohammad Bakhtiar Rana	

I hereby declare that the work submitted is my own work. I understand that plagiarism is defined as presenting someone else's work as one's own without crediting the original source. I am aware that plagiarism is a serious offense, and that anyone committing it is liable to academic sanctions.

Rules regarding Disciplinary Measures towards Students at Aalborg University:
http://www.plagiarism.aau.dk/Rules+and+Regulations/

Date and signature(s):

June 2nd, 2021

Signature and date

Abstract

The current slow adaptation of Big Data (BD) in internationally operating small- and medium sized businesses (IOSMEs) has been identified as a problem of economic importance. By analyzing the challenges and achievable benefits of BD adoption, this thesis will look into beneficiary options that support such a implementation with regards to the limited capabilities of IOSMEs. Further, through a comparative case study of external service providers that enable IOSMEs to conduct Big Data Analytics (BDA) their relation towards overcoming these challenges will be examined. With investigating BD through the Resource Based View and the Resource Dependency Theory in combination with the insights from the empirical analysis, it was established that a sustainable competitive advantage through the utilization of external service providers cannot be realized. The critical point in achieving such a competitive advantage proved to be the build-up of internal BDA capabilities that inherit the properties of the VRIO framework. However, the utilization of external service providers supports IOSMEs in overcoming challenges when adopting BD, which makes it a viable option in the early adaptation phase.

Keywords: Big Data, internationally operating SME, BDA, Resource based view, external service providers

Acknowledgements

I would like to express my biggest gratitude towards my professor and supervisor Mohammad Bakhtiar Rana. His valuable guidance helped and challenged me towards the realization of this thesis. I hope it pays a little tribute towards my Austrian academic ancestors.

As this master thesis is a milestone in my academic career, I reflect back towards other challenges I overcame and the support I received along the way. A major encouragement throughout my life was my family who always helped me in my advances, no matter the distance or difficulty. Then there are my friends who have fully accepted my lack in social contact, for which I cherish them even more. Lastly, I would like to thank my wonderful girlfriend. I cannot express in words how grateful I am to have you by my side.

It is thanks to you all I managed to be where I am now.

Weit hamas bracht - danke für alles!

Table of Content

List of figures

List of tables

1. Introduction

Small-and-Medium-Enterprises (SMEs) play a vital role in our economy as they represent 99% of all businesses and generate 60% of employment in the OECD area (OECD, 2019). While Big Data (BD) opens up new possibilities for the improvement of business activities, it is only slowly adopted by SMEs. Internationally operating SMEs (IOSMEs) have an even higher potential for advancing their business with BD, as more information is available to them which provides a greater leverage to benefit from it. Still, the complexity of integration due to various factors proves to be a too-high barrier for adopting BD in such companies (European Commission, 2019).

Previous research has mostly disregarded the option of utilizing third party solutions to integrate BD in smaller companies due to high costs, loss of control and dependability (Coleman et al, 2016; Dittert et al, 2017; Rajabion, 2018). But with the technological advancements made in recent years, the prices for computing power and data storage are getting more and more feasible. This also translates into more reasonable costs when taking advantage of external BD services. This raises the question if such services are still not a plausible solution for BD integration in IOSMEs. Thus, the research question (RQ) for this thesis is stated as

How can internationally operating SMEs successfully integrate Big Data with solutions from external providers?

As BD research is still in its infancy, the application of BD in smaller companies has been not sufficiently examined. This research gap gets even more prominent in combination with BD and the internationalization efforts of IOSMEs. Additionally, literature concerning the utilization of third-party service providers that help integrating BD in smaller companies has not been sufficiently examined as well. The reluctant academic process in this field is stated as a major contributor to the slow adaptation of BD (Iqbal et al, 2018). This thesis will examine solutions to overcome the barriers of BD adaptation with the help of third parties in an effort to bridge these two gaps in research and support the progress of BD utilization in IOSMEs.

2. Problem Formulation

The purpose of this paper is to obtain a critical evaluation of how IOSMEs can adopt BD in their business practices with the help of external services. This will be examined through an extensive literature review and an empirical analysis, where the following research topics (RT) will be researched, identified, and analyzed in order to answer the overall RQ:

RT1: What are the benefits of BD implementation and how does it lead to competitive advantage?

The first RT will examine the benefits when adopting BD in IOSMEs and how it can lead to competitive advantage. Applying the Resource Based View (RBV) theory, together with the VRIO framework will provide an understanding of the successful BD integration.

RT2: What are the challenges of BD implementation and how do they lead to the slow adaptation of BD?

Examining this RT will lead to a general understanding of the factors which oppose BD adoption. The potential challenges can occur internally as well as externally and will be divided accordingly. This supports an understanding of the general slow adoption of BD practices in IOSMEs. Additionally, it can be outlined which factors enable a successful implementation.

RT3: What are the options for BD implementation in IOSMEs?

IOSMEs have access to a multitude of different technical tools that support a successful integration of BD with regards to their limited resources. These suitable internal and external options for BD adaptation in IOSMEs will then be examined towards their implications on benefits and challenges. The Resource Dependency Theory will further help understand how external solutions can lead to a successful integration of BD.

RT4: Are there solutions from external providers for BD implementation that are aligned with the needs of IOSMEs?

The initial literature search has shown that the possibility of utilizing services from external providers for BD integration has been mostly disregarded in the literature. An empirical analysis shall investigate external possibilities for IOSMEs which are aligned with their limited resources and capabilities. Further, the analysis will show if these solutions can help IOSMEs overcome the challenges to fully exploit the benefits of BD.

The literature examined in this project will be in the field of business and information technology and subsequently be examined through a literature review. By combining the findings with the empirical analysis, this project will support the underdeveloped field of academic research in this area.

3. Definition of Key Concepts

The key areas from the literature review will be explained in the upcoming part to provide a basis of understanding. These concepts will be explained in the context of the problem formulation.

3.1 Definition of internationally operating SMEs (IOSMEs)

SMEs can be described as independent firms with a varying number of employees between different countries. The categorization in SMEs is segregated by number of employees and by financial assets. Small enterprises with a range of 10 to 49 employees should not reach a turnover higher than EUR 10 million, while medium sized enterprises between 50 and 249 employees should not exceed a turnover of EUR 50 million (OECD, 2005). In terms of balance sheets, small businesses have a limit of EUR 10 million, while medium-sized enterprises should not exceed EUR 43 million (OECD, 2005). In this thesis, IOSMEs are defined as SMEs that operate across national borders. Further, the researcher argues that findings within the literature directed towards SMEs are applicable towards IOSMEs as well, as they share the same limited capabilities. This consideration is further reassured through the already underdeveloped research of BD implementation in SMEs, which only worsens in the context of IOSMEs. In order to comply with scientific writing, the literature review will use these acronyms according to the examined articles.

3.2 Big Data (BD)

BD is a buzzword and often used as an umbrella term for multiple functions and processes. As mentioned by multiple authors, BD is sometimes mistaken for analytics (Ferraris et al, 2019) and the interconnectedness of BD and other related concepts lead to confusion (Provost & Fawcett, 2013). The later conducted literature review stands true to this fact, as many articles mentioning BD are switching fluently between the definitions of BD and Big Data Analytics (BDA). This requires a clear separation in order for this thesis to not contribute further to the confusion.

The Oxford definition defines BD as *"sets of information that are too large or too complex to handle, analyse or use with standard methods"* (BD, 2021). Nevertheless, some authors claim that BD is not only about the data volume. Russom (2011) states that although data volume is the primary attribute, the variety of data collected by different platforms (social media, business processes, geospatial) and the velocity (frequency, exponential growth of data) further contribute to its size. Recent academic papers state an exponential increase of volume in data every day and its fast-moving advances in capitalizing it, which further argue for this definition (Iqbal et al, 2018). Throughout this thesis, BD is seen as a resource comprising vast volumes of data.

3.3 Big Data Analytics (BDA)

Runkler (2016; p.2) defines BDA as *"[…] the application of computer systems to the analysis of large data sets for the support of decisions."*. It includes the processing, analysis, model creation and model usage of data in order to support decision-making (Noonpakdee et al, 2018). Another definition for BDA explains it as the process of extracting, generating, interpreting, and categorizing useful information through the compression of an enormous amount of data (Bertello et al, 2020). Ferraris et al (2019; p. 1924) defines it as *"the process needed to comprehend the conglomerate of data in order to extract and generate useful information and knowledge which, through interpretation and categorisation, lead to more effective management"* (Ferraris et al, 2019; p. 1924). Conclusively, BDA is the company's capacity to manage, process and analyze BD.

4. Methodology

This chapter discusses the positioning of this paper within the theory of science. The specific areas of the philosophy of science will be explained through the works of Burell and Morgan (1979) and Arbnor and Bjerke (2009). These authors provide the necessary tools to further explain the methodology used for this paper.

The aim of the methodology is to explain the methodological approach through which the research questions will be answered. To explain the literature review, the framework of Arbnor and Bjerke (2009) will be used as shown in Figure 6. The review methodology will include ultimate presumptions, paradigm, methodological view, operative paradigm, and study area according to Figure 1. The operative paradigm will include a review protocol, synthesis, and a literature review model. The theory of science is concerned with the fundamental assumption

about what knowledge can be defined as and how it can be utilized to obtain an understanding of reality within the review and comprises the paradigm, ontology and epistemology of this thesis.

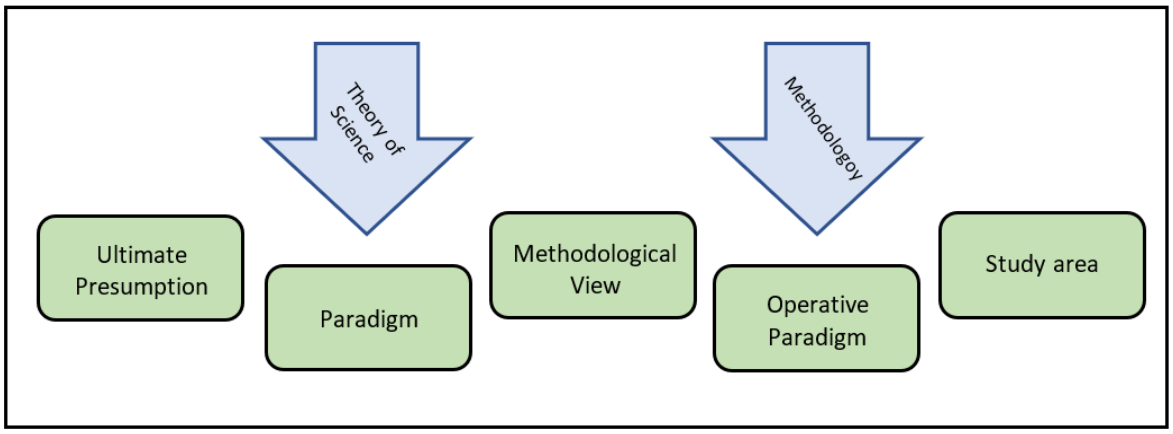


Figure 1: Methodological approach

Source: own compilation based on Arbnor & Bjerke (2009)

4.1 Ultimate presumptions

Before conducting a study the ultimate presumptions have to be laid down. *"These presumptions differ between views, and the different views, therefore, present different ways to understand, explain and improve"* (Arbnor and Bjerke, 2009; p. 4). Accordingly, the perceived reality of the researcher guides the study to the made presumptions. If the research does not reflect on which particular view forms the understanding of the investigated topic, it can lead to the blocking of observations, understanding and arguments (Arbnor and Bjerke, 2009).

4.2 Paradigm

The paradigm is laid down by the ultimate presumptions. The main distinction in research approaches, as defined by Burell and Morgan (1979) is the subjective (Interpretive) and objective (Functionalist) dimension with their respective paradigms. Figure 2 illustrates the different dimensions of ontology, epistemology, human nature, and methodology. These integral parts will be explained in detail in the following chapters. The main distinction of these paradigms is concerned with the view of reality.

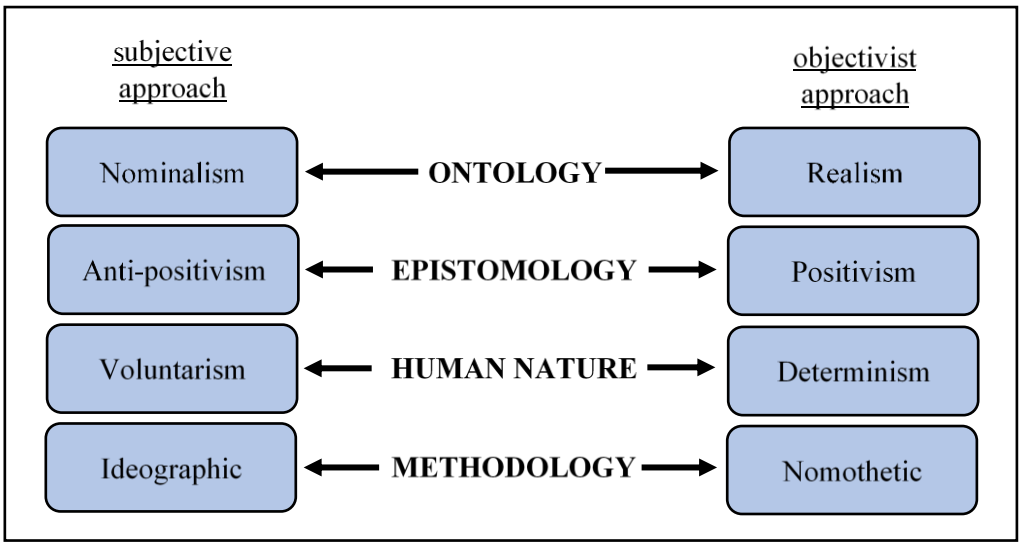


Figure 2: Subjective-objective dimensions

Source: own compilation, based on Burell & Morgan (1979)

In general, the two dimensions each have their own methodology. In subjective approaches, these methods are centered around personal, qualitative data (Ideographic), such as interviews or open questions in surveys. The objective approaches are more engaged with statistics or surveys with closed answers.

4.2.1 Ontology

Ontology tries to answer whether reality is independent of the individual or the product of individual consciousness. The work of Burell and Morgan (1979) separates the perception of ontology into two different views: Nominalism and Realism. Nominalism defines reality as subjective and as a result of dynamic interactions between individuals and the environment. As such, reality would be based on the context and the perception of individuals and not what is universal and general. Realism, on the other hand, defines reality as a non-socially constructed system. Thus, reality is independent from external factors and only one independent reality exists.

4.2.2 Epistemology

A mnemonic to describe Epistemology is how the seeker of knowledge acknowledges knowledge. It describes assumptions about the foundation of knowledge or how one understands the world and communicates knowledge to others. Following again the explanation

of Burell and Morgan (1979), epistemology can be defined in two different dimensions: Positivistic, related to subjectivism, and Anti-Positivist, related to objectivism.

In Positivism, knowledge is perceived as a cumulative process, where new insights contribute to existing knowledge and false hypotheses are disregarded. The researcher takes the role of an external observer. As such, the researcher objectively searches for regularities and relationships between elements of investigation, while not being involved in the social context of the researched system (Burell and Morgan, 1979). Anti-positivists see the world as relativistic and only possible to be understood through the view of individuals that are active in the field of study. This view tends to reject the possibility of objective knowledge as only through *"occupying the frame of reference of the participant in action"* (Burell and Morgan, 1979; p. 5) understanding can take place. As such the researcher has to be part of a socially constructed system in order to process and create knowledge.

4.2.3 Human nature

Burell & Morgan (1979) describe human nature as concerned with the relationship between human beings and their environment. Determinism sees human beings and their experiences as products conditioned by their environment and external circumstances. On the other side of the spectrum, Voluntarism sees humans as beings with free will, which control rather than being controlled by external factors. Here, humans and their actions are regarded as the masters of their environment.

4.2.4 Methodology

The methodology is separated between the Ideographic and Nomothetic view. The Ideographic approach only understands the social world by extracting knowledge from the examined subject. It emphasizes the need for detailed, subjective accounts in order to unfold the characteristics and nature of the subject under investigation. On the other hand, the Nomothetic approach stresses the need for basing research on systematic protocols and techniques. It utilizes quantitative techniques such as surveys, questionnaires, and standardized research instruments.

4.3 Methodological View

Arbnor & Bjerke (2009; p. 423) define the methodological view as a *"consistent set-up of ultimate presumptions, concepts and principles guiding creation of knowledge"*. A methodological view that fits the chosen type of review needs to be selected.

Arbnor and Bjerke (2009) distinguish three different views, as seen in Figure 3. There, the methodological approaches, paradigmatic domains, and the distinction between objective and subjective approaches are illustrated.

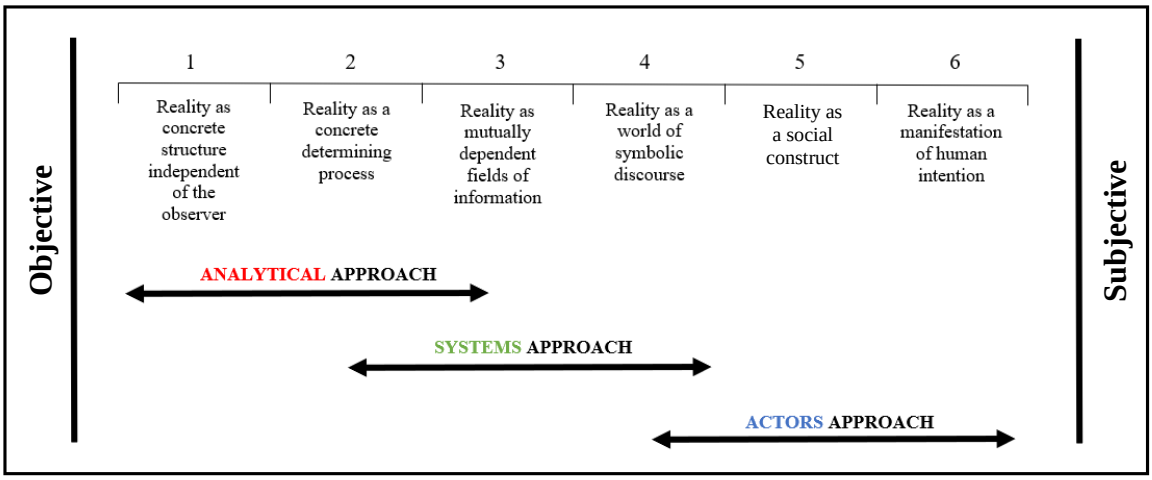


Figure 3: Methodological approaches and paradigmatic domains

Source: own compilation, based on Arbnor & Bjerke (2009)

Within the Analytical approach knowledge is created independent from the observer, e.g. with the possibility of subjective experience and assumptions. Arbnor & Bjerke (2009) further explain the analytical approach to be based on facts with a summative character, meaning once the research understands the different parts, they can be put together to understand the phenomenon from a holistic perspective. The Systems approach combines both qualitative and quantitative data and sees reality as mutually dependent fields of information. The actors approach defines the view of reality as a social construct and a manifestation of human intention and is linked to a subjective paradigm (Arbnor & Bjerke, 2009).

4.4. Choice in Paradigm and Methodological View

The aim of this thesis is to investigate how IOSMEs can successfully integrate BD with the help of external service providers. Initially, the aim is to find the benefits of BD together with potential challenges IOSMEs can encounter on the way. This will lay the ground for

understanding the topic which will then be examined together with well-known existing theories within international business such as the RBV, competitive advantage, the VRIO framework and the resource dependency theory. As there is a gap within the literature of how these aspects can affect IOSMEs, this thesis will aim to provide additional information on these aspects. Lastly, the suitable options for IOSMEs will be investigated and further filling another gap within the literature by illustrating the available options of external providers for IOSMEs by conducting empirical research based on secondary data. This thesis will pursue an objective approach with a Functionalist paradigm based on the proposition of Burrell and Morgan (1979) based on the following justification.

The objective approach has been found to be the most relevant for this thesis and will thus guide the other aspects within the methodology. Investigating the topic of BD implementation on IOSMEs is explaining a process rather than an individualist viewpoint as it would be the case within a subjective paradigm. Further, within an objective paradigm, the aforementioned challenges and benefits are existing outside of an individual's consciousness and are not subject to interpretation. The same applies to the available options and external solutions to implement BD. BD is categorized as a resource which a company can either implement or not and can be seen as a non-social phenomenon. Therefore, the decision is to follow an objective approach.

The ontological position chosen for this thesis is Realism, following the objective paradigm. Nominalism is not suitable, as the overall research question is not concerned with individual's perceptions but rather with which external options for BD implementation exist for IOSMEs and how they can be successfully integrated. This lays within the concept of universal truth. Therefore, this viewpoint is aligned with the realistic view which can be understood as a non-socially constructed system and can thus be rationally examined. Overall, the problem formulation is of pragmatic nature, and the thesis aims to provide a pragmatic solution for a pragmatic problem.

Followed by the chosen ontological position of Realism, the chosen Epistemology in this thesis is Positivism. The aim is to generate new insights through the empirical analysis which will comprise two case studies with the researcher being an observer rather than involved in a socially constructed system, as it would be the case in an Anti-Positivist Epistemology. The slow adaptation of BD in IOSMEs is factual and can be objectively observed. It is not necessary

for the researcher to be personally involved to understand the phenomenon to find the most suitable solutions for IOSMEs.

Regarding the dimension of human nature, the researcher chooses to follow the view of Determinism. Managers in IOSMEs can express their 'free will' in choosing between implementing BD or not, which could be interpreted to be within the Voluntarism view. However, the implications governing a successful implementation are determined by their environment. Further, choosing not to implement BD can hurt the survivability of the IOSME due to competitors implementing it before them. Thus, the involved individual's actions are determined by the external environment, which justifies the deterministic view used in this thesis.

The methodological debate between Ideographic and Nomothetic is concerned about the researchers personal involvement to understand the phenomenon in an every-day setting. As mentioned above, given that this thesis is answering a very pragmatic problem which does not require the researcher to be involved, the chosen methodology is Nomothetic, which is in line with the overall Functionalist paradigm.

In terms of methodological view for this project, the analytical approach is the most suitable as it fits the objective paradigm and the chosen ontological and epistemological views, together with the use of secondary datasets. Within the Analytical approach knowledge is created independent from the observer, and the focus of this thesis lies on explaining the findings and adding knowledge to existing literature. The researcher aims to understand all parts that are crucial for IOSMEs when integrating BD into their business practices. Therefore, the researcher will examine the different parts that are integral for answering the research question, such as how BD works, what the benefits and challenges are and what options exist for implementation. By understanding the individual parts, they can ultimately be added together to understand the whole picture relevant for IOSMEs and how these insights can be translated into a successful integration. The Actors view aims to understand human intentions within a social construct in depth by using qualitative research methods, while the Systems view favors a mixed-method approach. Both are irrelevant for this research question because the focus lays not on the motivation to implement BD but rather on externally governed factors for a successful implementation.

4.5 Operative Paradigm

Arbnor & Bjerke (2009; p. 43) define an operative paradigm which *"relates a methodological view to a specific study area"*. The purpose of it is to show the underlying connection between the basic assumption, the applied paradigm, methodology and the area of research. It explains the process of obtaining knowledge about the area of inquiry. In this section, the purpose lies in describing how knowledge is obtained and how the empirical data will be collected and assessed.

Bryman & Bell (2015) explain that researchers need to distinguish between research design and research methods. Research design is about how to collect and analyze data which connects research problems, theories and methods as well as results of a research. On the other hand, research method is a technique that is concerned with data collection in order to answer the overall research question. Researchers have different methods at their disposal to answer the research question which then has to be choose between to find the most suitable to investigate the phenomenon.

4.5.1 Research Design and Research Method

The Research Design and the Research Method correspond to the Operative Paradigm proposed by Arbnor & Bjerke (2009). The outline of this thesis is visualized in Figure 4.

Prior to the secondary data collection, a thorough literature review is conducted. With the literature review, the first three research topics will be answered, to understand what has already been written on the investigated issue. Following this, a conceptual framework will be developed to summarize the main findings from the literature, including the most encountered topics and issues related to the overall research question. The conceptual framework shall act as an analytical tool to summarize the ideas found in previous research. The conceptual framework will then be tested by the findings within the empirical data to extend existing knowledge in this study area. This is, again, aligned with the objective approach and the analytical view which states that reality is constructed independently of individuals. Further, it can thus be concluded that this thesis follows a deductive approach.

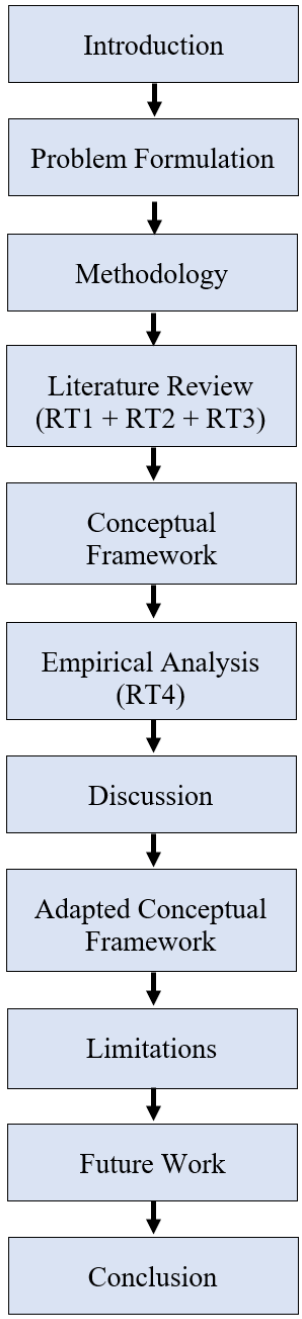


Figure 4: Thesis outline

own compilation

The deductive approach can be explained as theory guiding the research, showing the relationship between theory and empirical research (Bell, Bryman & Harley, 2018). By following a deductive approach, the researcher will develop a conceptual framework which is based on previously written literature which will afterwards be tested using empirical data (Saunders et al., 2009). The deductive approach is mostly connected to hypothesis creation; however, this thesis will follow a non-hypothesis approach. This is further justified as hypothesis building requires measurable variables. This is not applicable for this thesis which focuses on a more generalized perspective towards successful BD implementation. Further, specific challenges are simply not measurable in numbers and can thus not be evaluated. Thus, no a-priory hypothesis will be stated in this thesis. Nevertheless, by sequencing the overall RQ into three RTs, the research will be simply linear, creating a logical and clear path to follow for the researcher and the reader.

Further, Bryman & Bell (2015) distinguish research design into three categories: experimental design, cross-sectional design, longitudinal design, case study design or a comparative design. The research conducted in this thesis is following a comparative case study research design with secondary data. The comparative case study design requires the same methods for contrasting cases. This allows the researcher to understand the phenomenon in a broader sense when comparing two contradicting cases. The selection of cases shall have the aim to provide an opportunity to learn from them and thus the selection of the cases should be thought through to provide the biggest benefit for research.

4.5.2 Data Collection Method

The first question the data collection is concerned with whether to gather primary or secondary data for this thesis. For the case of the stated RQ, there are numerous arguments that speak against a collection of primary data. The slow adaptation of BD practices in IOSMEs already

limits the available pool of possible candidates. This pool gets even smaller when extending the search towards IOSMEs. Further, as BDA capabilities are a source for competitive advantage, IOSMEs could be unwilling to share their experiences freely. As the RQ has a generalized format it would be necessary to obtain a multitude of companies to establish a level of validity to the findings. Such a process is an expensive and time-consuming undertaking that stretches far beyond the available resources for this thesis's researcher. Lastly, the ongoing COVID-19 pandemic has led to radical organizational changes that range from extensive layoffs to employees working from home. Arguably, this negatively impacts the willingness of IOSMEs to cooperate in academic research to due to more pressing, critical organizational matters. Due to these reasons, the researcher decided for the utilization of secondary data collection. This is in line with the comparative case study design, as it can be based entirely on secondary data analysis. However, the comparative approach is in need for suitable strong evidence in order to draw appropriate conclusions from the cases. Thus, the validity of this thesis has to be questioned. With the cases stemming from the external service providers' website, the researcher has to critically assess the gained information towards biases. Therefore, the addition of academic articles to critically assess the steps taken in the BD integration process was chosen in order to enhance the validity of the used cases.

This concludes that the collected data will be based on qualitative research. Qualitative research can be defined as *"a research strategy that usually emphasizes words rather than quantification in the collection and analysis of data"* (Bell & Bryman, 2011; p. 386). Research methods used within qualitative research can include participant observation, interviews, focus groups, or a collection and qualitative analysis of texts and documents (Bell & Bryman, 2011). Given that the empirical findings will be based on secondary data it can be concluded that the suitable research method is the collection and qualitative analysis of texts and documents, as no primary data collection will take place. As suggested by Bell and Bryman (2011), generally speaking, within qualitative research researchers initially start with broad outlines of concepts, which are then adjusted throughout the time of data collection.

In the next step the determinants for suitable companies were constructed. The initial literature review suggested that external services are not customized towards the needs of IOSMEs as such services are mostly provided by larger companies. Thus, the first criteria were for the external service to be an IOSME themselves, inhabiting the knowledge of crucial business aspects in IOSME customers. The second criterion was for the external service to provide a BD

infrastructure for their customers, as the construction of said BD infrastructure was related to the challenge of high upfront cost if an IOSME was to construct it on-site. Lastly, the case company has to utilize a optimal solutions for BD integration related to the findings in the literature review.

The cases were searched for by using online search engines. Primarily, the found optimal options were used as keywords in combination with abbreviations for IOSMEs. After finding possible cases, they were screened according to business features of IOSMEs, namely their employee count and turnover with the help of business community websites such as Owler. Lastly, through examining the companies' websites and related news articles it was established if they conduct business internationally in order for them to fall into the category of IOSMEs. This process resulted in two found companies that fit the criteria and were further assessed in the empirical analysis.

To construct the case data, each company website was searched for information about their underlying business processes. Such found cases were further enriched with news articles in order to prepare them for the data analysis.

4.5.3 Data Analysis

To analyze the qualitative data collected, this thesis uses Grounded Theory developed by Glaser & Strauss (1999). By systematically obtaining data and empirically developing a theory, it can be seen as a more systematic approach. Instead of creating a-priori hypotheses, it aims to examine the concepts that emerge from the data. This requires that the researcher has no preconceived theoretical ideas before starting the research. Within Grounded Theory, there are two distinctive approaches: the systematic, rigid Straussian approach and the open Glaserian approach. The main distinction between these two approaches revolves around that the Straussian approach with its prescriptive steps 'tortures' the data until it fits, whereas the Glaserian approach allows the data to speak for itself (Bell & Bryman, 2011) The Straussian approach was chosen as it provides the researcher with helpful ready-made tools for the analysis.

In the first step of the analytical process, relevant data according to the RQ is collected. The next step is the coding process of the data. Through the coding the need for new data will arise, which results in the need to sample theoretically. Thus, there is a constant movement between

the steps of data collection, coding and theoretical sampling. The theoretical sampling is concerned with selecting samples that refine and extend the description of the examined phenomenon. Thus, the sample size is not defined before the investigation but through the theoretical saturation. When the point of theoretical saturation is reached and no new is data is emerging from the data collection, the process of data collection is finished. Lastly, a theory is developed based on the empirical data. The analytical process of the Grounded Theory can be seen in Figure 5.

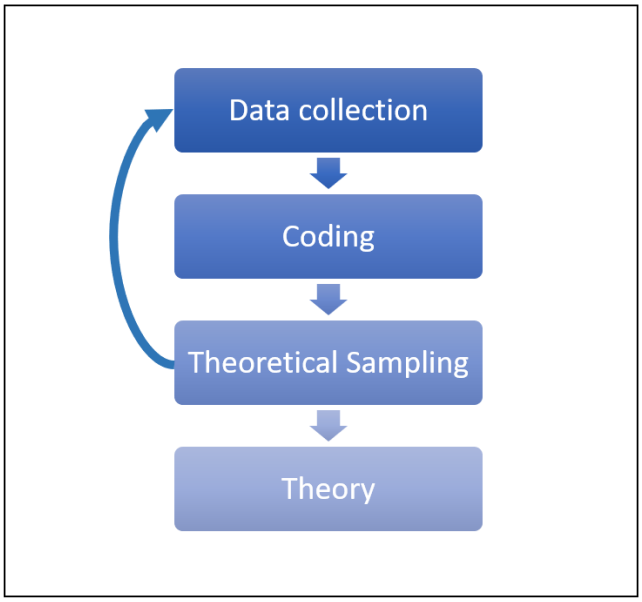


Figure 5: Analytical Process in Grounded Theory
Adapted from Bell et al (2019)

At the heart of Grounded Theory stands the constant comparison. This is the process of comparing the empirical data with each other in the analysis process to find similarities and differences between them (Bell & Bryman, 2011).

The data is coded through three steps, namely open coding, axial coding and selective coding. In the first step, the data is sequenced, analyzed, compared and categorized to provide an initial understanding of the data. The aim is to create fitting categories for the data, albeit just provisional. As the open coding relied on the previously investigated categories from the literature review, these categories will then be transformed in the step of axial coding (Bell & Bryman, 2011). The framework of the axial coding process is visualized in Figure 6.

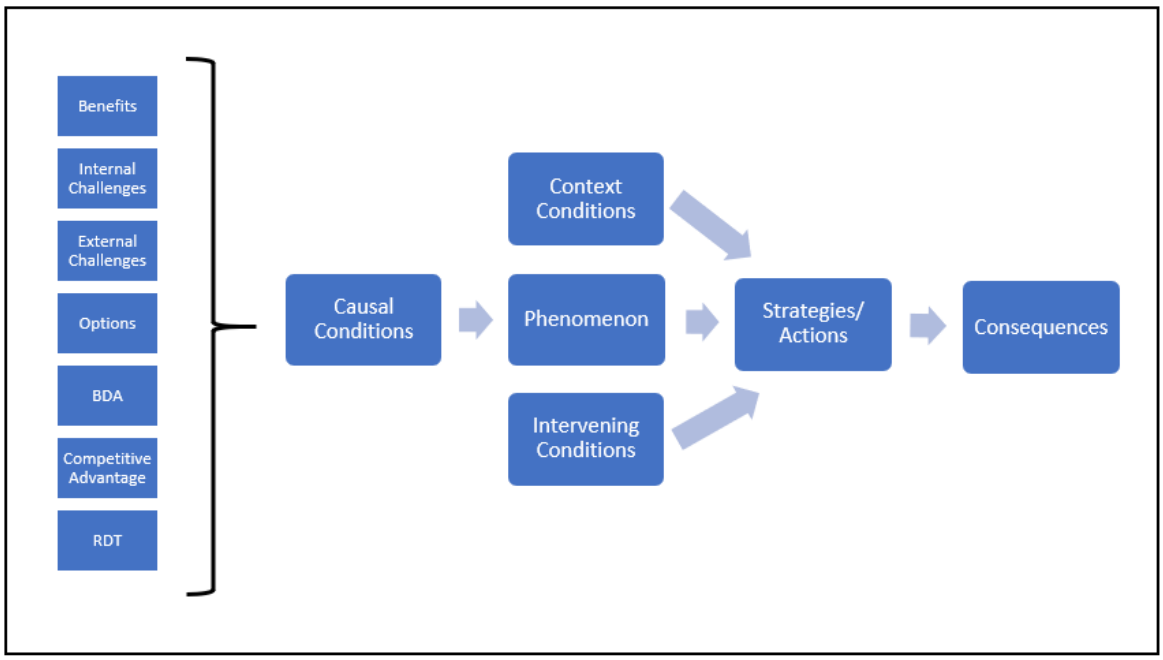


Figure 6: Axial coding framework
Adapted from Corbin & Strauss (1990)

In this framework, the Phenomenon takes the central part as the subject under investigation. The reasons that led to this Phenomenon are described in the Causal Conditions while the Strategies and Actions are the measures taken because of it. The Context Conditions describe the circumstances in which the Strategies take place and define the context of the phenomenon. Lastly, the Intervening conditions describe the attributes that influence the Strategies and Actions. Through this, the categories are individually examined in-depth to establish relationships and explanations between them. The last step refines the analysis towards a core category to which the other categories can be linked. Here, the initial relationships are tested to provide further validity.

The Grounded Theory focuses on inductive analysis (Chun Tie et al, 2019). Nevertheless, there is criticism in academic research arguing for using a deductive qualitative analysis within Grounded Theory (Gilgun, 2013). Gilgun (2013) introduced thus the Deductive Qualitative Analysis which ensures the connection of theoretical concepts and data, which ultimately allows the researcher to reject theories and concepts that are not suitable. She further argues that the procedures of grounded theory can be conducted in a deductive manner, as long as the researcher enters the research area with an open mind, which this thesis will ultimately follow.

4.5.4 Case Company descriptions

Two companies will be examined in the empirical analysis. Both of them fit the description of an IOSME and are engaging in BDA activities.

Transmetrics

Transmetrics was founded in 2013 in Sofia, Bulgaria. With 35 employees, they offer Software-as-a-Service for cargo transport and logistics service providers by applying artificial intelligence, Data Mining and predictive analytics. They conduct data cleansing, demand forecasting and predictive optimization as the leading software provider in the logistics industry. (Company Transmetrics, 2021)

Since the rise of e-commerce in 2013, cargo transporting shifted from containers and pallets towards boxes and envelopes for transporting goods. These smaller shipments are subject to fixed departure networks from logistic companies in order to group them together, which often results in empty spaces in collective shipments. To give an example, a container transport operation is on average 24% empty and for parcel networks, such as DHL or FedEx, empty spaces go up to 60%. Here, Transmetrics offers a solution by utilizing cloud computing services to obtain logistics information from its customers, analyze it and offer various solutions to enhance efficiency and lower carbon emissions in the process. In detail, they optimize transport planning by predicting future shipping volumes up to 6 weeks in advance in order to enable their logistics customers to fill empty spaces, thus improving their capacity utilization. Transmetrics states in one case a reduction from 43% of empty space to 18%, which in turn decreases the amount of traveling vehicles by a further 20%. Their service is offered at a rate between EUR 3,500 to 5,000, depending on fleet magnitude and other logistics-related measures. (Transmetrics, 2021)

With the current COVID-19 pandemic positively affecting Transmetric's business, they opened a new commercial headquarter in Amsterdam, Netherlands in February 2021. Now being an IOSME, they plan to capitalize on the changes in logistics in the European market due to the UK leaving the EU. With a growing customer base in Western Europe, Netherland's importance to trade and a vast tech-talent pool, they hope to further grow their business. (Transmetrics in Amsterdam, 2021)

AtScale

AtScale was founded in 2013 in San Mateo, USA ("AtScale Company", 2021). Currently, they employ 125 employees ("AtScale Profile", 2021). They offer virtualization of data silos, meaning they bundle company data on a single cloud-platform in order to break up existing silos of knowledge in companies ("AtScale Company", 2021). Further, they offer to modernize the existing BDA architecture of companies and transform data through autonomous data engineering in order to make it ready for analysis. Through the deployment of these functions, they enable companies to develop business decisions based on BDA faster and more accurately. Their mission is to create an enterprise shared data intellect for their customers ("AtScale Company", 2021).

They benefit from the slow adaptation and regressive culture towards BDA from IOSMEs. Many companies are overwhelmed with the challenges from BD and its 5V's. Here, AtScale offers an all-inclusive service to enable companies to take the leap for BDA integration. Further, they offer a modernization to existing practiced analytics already established in companies by enabling the data to be analyzed by common and unfavorably tools for BDA such as Excel ("AtScale on Hadoop", 2021). Through that, existing analytics practices do not need to be changed, as the used tools are compatible with the new infrastructure.

By 2018 they opened a second headquarters in Boston, US and a developmental center in Sofia, Bulgaria, now fulfilling the requirements for being an IOSME ("AtScale Company", 2021). The CEO of AtScale, Chris Lynch, stated that the reason for the establishing a second headquarter was due to more cost-effective human capital with a principal engineer being about 70% cheaper at the new location ("AtScale CEO Interview", 2021). Additionally, the new time zone allows for a better cooperation with their development center in Bulgaria ("AtScale CEO Interview", 2021).

4.6 Literature Review Methodology

The following chapter is concerned with the review protocol and shall illustrate how the research question is intended to be researched. This protocol includes how the literature will be synthesized and the study quality and assessment.

4.6.1 Review protocol

This literature review will use a narrative synthesis which is concerned with understanding rather than the accumulating knowledge. *"The literature review is for them a means of gaining an initial impression of the topic area that they intend to understand through their research. Narrative reviews therefore tend to be less focused and more wide-ranging in scope than systematic reviews. They are also less explicit about the criteria for exclusion or inclusion of studies"* (Bell & Bryman, 2011; p. 101). As Lisy and Porritt state *"it enables investigation of similarities and differences between studies, exploration of relationships within the data and assessment of the strength of the evidence, and results in a summary of knowledge related to a specific review question that may be used to inform practice or policy."* (Lisy & Porritt, 2016; p. 201). This concludes that within the literature review both qualitative and quantitative research will be taken into consideration to provide a holistic view of the overall topic. The initial step aims to provide an understanding of key indicators influencing the decision to incorporate BDA within IOSMEs and the benefits that can be gained. The second search will examine the possible options to successfully implementing BD with respect to the found key indicators. Through a narrative synthesis similarities and differences between researchers' theories will be examined and a summary of knowledge can be developed. After conducting a framework based on these findings, said framework will then be tested through the empirical findings.

4.6.2 Study quality and assessment

Validity, replication and reliability form the most relevant criteria for the evaluation of business research. Validity is concerned with the integrity of conclusions that are generated from a study. The ability to make studies capable of reaching the same conclusions is withheld in Reliability. Lastly, Replication is implying the capability of replicating a conducted study (Bell & Bryman, 2011).

For this thesis, the validity is fulfilled through the selection of academic books and articles which are written by field experts, confirmed by the reference list in each paper. In terms of reliability, this paper relies on the background of the author and prior published books and articles. The Operative Paradigm with its explanation of procedures ensures the replication.

5. Literature review

Before the stated RTs can be answered, it is necessary to explain the concept of BD in detail in order to form a prior common understanding for the used terms and the related concepts.

5.1 Big Data

BD is often reclaimed as the oil of the 21st century. Academic articles already proclaim a *"gold-rush-like atmosphere"* (Espinosa et al, 2016; p. 1) and see BD as the most significant technological disruption in business and academia since the advance of the Internet (Agarwal & Dhar, 2014). The rise of BD in combination with the advancing datafication of our society has led to unimaginable vast flows of data. According to Coleman et al (2016), the daily generated amount of data equals to 2.5 x 10^6 terabyte. To put it in perspective: If one would copy this daily amount of data onto regular-sized USB-sticks with a memory of 8GB and lay them next to each other, the combined length of these USB sticks would span the world 1.5 times. With the information in the world doubling every 20 months, businesses are increasingly compelled to find measures to capitalize it (Runkler, 2016). BD's economic importance is highlighted in its market size, depicted in Figure 7.

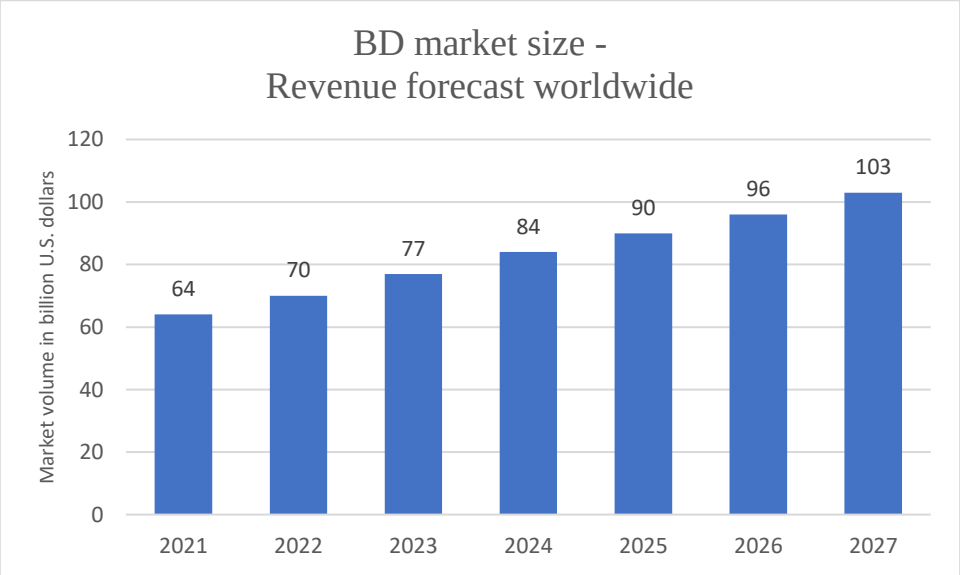


Figure 7: BD market size forecast 2021 – 2027,

Source: own compilation, based on (Statista Research department, 2021)

Weiss & Indurkhya (1998) were the first academic reference citing BD in computer science. They focused on predictive analytics and the potential of BD leading to stronger conclusions.

This potential is due to the increased usage of digitalization, e.g. smartphones, the further advance of the internet and social networking (Ferraris et al, 2019). Next to the increased size of data, an additional reason for the rise of BD is the increasing availability of technological resources. Mbassegue et al (2016) argues that BD projects which support companies in their decision-making could have been already realized years ago as the computational processes were already available. But the required technological investment to facilitate such processes were too high for a feasible introduction of this technology into smaller businesses. Now, computing power can be realized at a fraction of that cost, which makes BD financially available for businesses of any size (Mbassegue et al, 2016).

5.1.1 Five V's

Overall, the most used explanation for BD in academic articles are the five V's. The five V's are namely Volume, Variety, Velocity, Veracity and Value (Ferraris et al, 2019). Depending on the publishing date of each reviewed article, the number of V's ranges between three and five. This is due to the growth in the areas of research and new fields of applicability in which BD can be utilized. The definition of each 'V' can be found in Table 1.

Table 1: 5 V's of Big Data

Volume	The Volume of data grows exponentially (Bertello et al, 2020). The storage capability of the entire internet from 20 years ago is now generated every second (Ferraris et al, 2019). Looking into the future, a study mentioned by Iqbal et al (2018) reports that 2.5 Exabyte of data is produced daily. To put it into perspective, one Exabyte equals one billion Gigabytes. As mentioned before, this size doubles every 20 months (Runkler et al, 2016). The increased amount of internet enabled devices plays a key role in this massive growth of data (Iqbal et al, 2018). A greater volume further relates to a greater potential of insights within the data, which enriches its application and exploitation (Mbassegue, et al, 2016).
Variety	Variety relates to the many sources BD derives from (Bertello et al, 2020). The different sources have an impact on the structural heterogeneity in a dataset, which can be structured, semi-structured and unstructured data (Noonpakdee et al, 2018). With more and more data being generated from various digital platforms, its structure gets increasingly diversified (Iqbal et al, 2018). Thus, BD can be obtained e.g. from messages, images, readings from sensors or GPS signals from mobile phones (Ferraris et al, 2019). This Variety in data is significantly leveraging the Value of data, as combining different sources leads to an increase in validity of the resulting insights (Dam et al, 2019).
Velocity	Velocity refers to the speed at which BD is generated (Noonpakdee et al, 2018). In the economic world a key factor to success is to be able to make decisions faster than your competition (Ferraris et al, 2019). Thus, BD requires prompt and faster responses in order to manage and analyze it (Bertello et al, 2020). Nowadays, BD is able to provide real-time analysis (Iqbal et al, 2018). With the increasing speed in which BD can be utilized, it is possible for companies to become increasingly faster and agile in their decision-making progress (Ferraris et al, 2019).
Veracity	Veracity is concerned with the trustworthiness of the collected data (Bertello et al, 2020). It refers to possible unrealizabilities or uncertainties in data sources (Noonpakdee et al, 2018). With the exponential growth of data, the risk of noise, incomplete or out-of-date datasets is increasing as well (Ferraris et al, 2019).
Value	Extracting economic benefit from BD is naturally the overall goal of adopting BD practices and thus of vital importance (Bertello et al, 2020). Value can be generated by accumulating data and selling it to other companies which then generate insights from it. If the available data is very specific and hard to access for other companies, its potential value increases. Another obvious choice is the possibility to support business decision-making and subsequently increase a company's competitive advantage (Noonpakdee et al, 2018).

These 5V's circumvent the business perspective of BD, which needs to be considered when adopting it into business practices. As each V poses different challenges and opportunities for an IOSME trying to benefit from it, it is the cornerstone for any integration process. The 5V's and their implications are visualized in Figure 8.

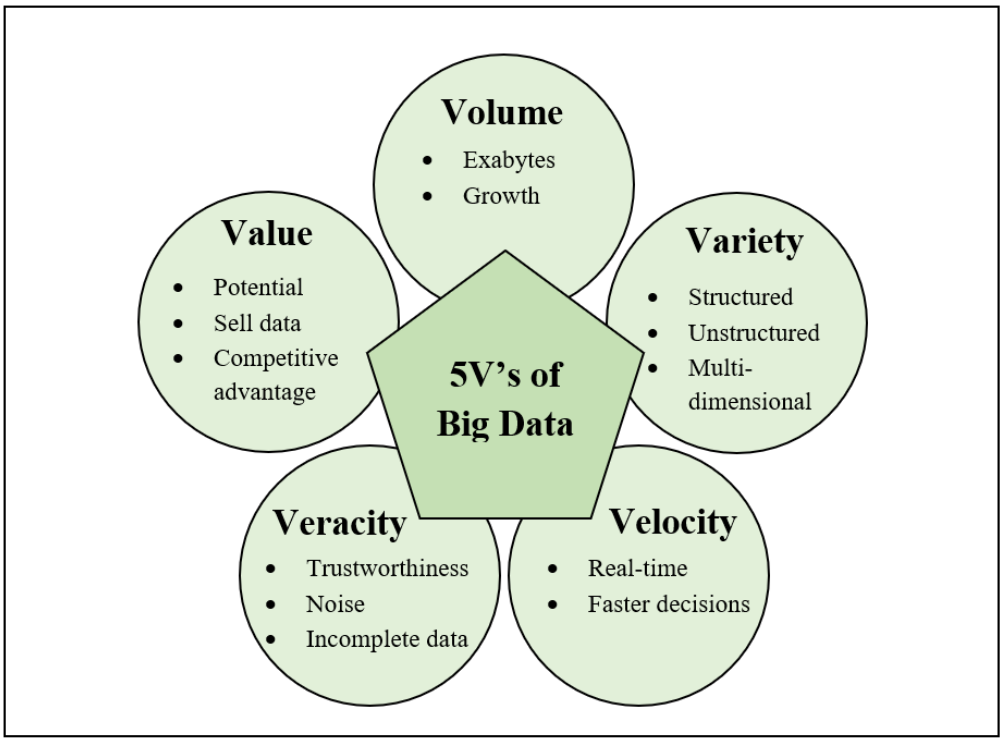


Figure 8: 5V's of Big Data

own compilation

5.2 Achieving benefits of BD implementation

In this section the mentioned benefits from the reviewed literature will be investigated. Next, the difference between BD and Big Data Analytics (BDA) will explained to understand how these benefits can be achieved. As BD is defined as a resource within this thesis, the topic is then explored in accordance with the RBV theory. This which is closely linked to competitive advantage and the VRIO framework.

5.2.1 Big Data Analytics

BDA is different from traditional data analytics, as it focuses on advanced analytical processes to detect real-time changes in e.g. purchasing decisions (Schuilling, 2020). Generally, BDA

consists of two things: BDA capabilities and BD infrastructure (Bertello et al, 2020). The BD infrastructure will be explained within the challenge of IT Infrastructure. Regarding BDA capabilities, they are the skills used in the analytics process and are concerned with the ability to acquire, store, clean, transform, analyze and interpret large amounts of diverse data which provides meaningful information and allows them to discover insights in a timely fashion (Mbassegue et al, 2016). The main objective of BDA is to develop actionable insights from BD which then are used to support strategic decision-making in order to positively affect business performance. This is achieved through the analytics process which covers all areas of BDA and can be seen in Figure 9. Table 2 illustrates each step of the process in detail.

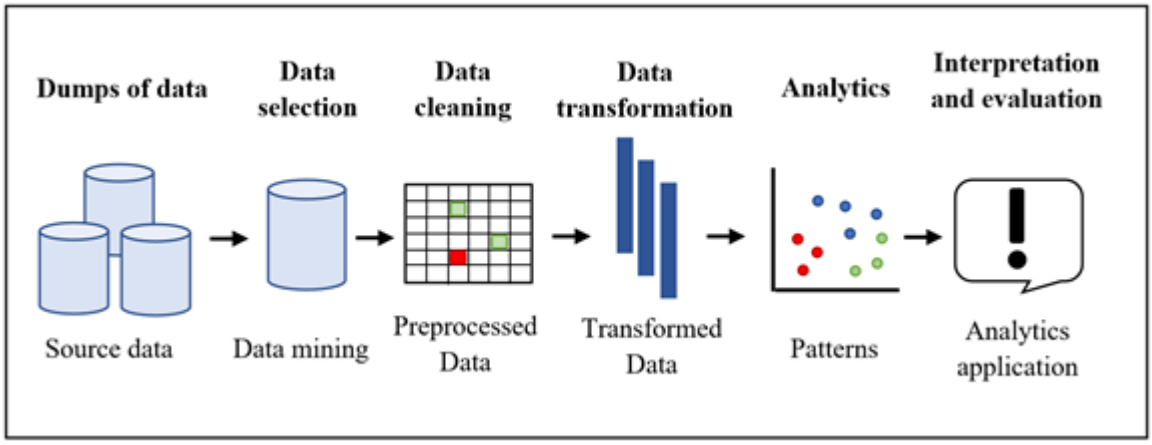


Figure 9: Analytics process

own compilation, based on (Mbassegue et al, 2016)

Table 2: Detailed Analytics Process

Data Acquisition	An IOSME can utilize internally generated data, use publicly available data or acquire data through a vendor which provides access to the required information (Mbassegue, et al, 2016). **Example:** Statistics Denmark provides publicly available datasets on different aspects of living conditions, consumption and so on. Such data is anonymized in order to not be in breach of Legal Regulations. When buying consumer data from vendors, the prices depend on the detail of information. Plain demographic information is available for lower costs than e.g. the web search history of consumers. **Example:** one can buy over 4 million consumer e-mail addresses from France for EUR 40 from such a vendor ("Buy Database", 2021). An IOSME has to keep Data Quality in mind → such a dataset does not automatically mean that its fully applicable. The Variety of the 5V's plays an important role as the gathered data can come in different formats. **Example:** Social media data is usually highly unstructured and contains noise. The noisiness of data can be understood as additional meaningless information which can highly affect the certainty with which it can be used. In more severe cases, the noisiness can lead to data corruption through which a user system cannot interpret data correctly.
Data Extraction	Data extraction is based on their importance to the analytical problem → Data Mining Data Mining is utilized to extract relevant data from dataset for the analytics process through which insights will be uncovered. It can also be referred to as 'KDD' – knowledge discovery in databases (Dittert et al, 2017). Data Mining is the challenging process of identifying valid, novel, potentially useful, and ultimately understandable relationships in data (Fayyad et al., 1996). → Data Mining itself is a process of analytics. Its application does not include the analysis of large datasets, as is the case in BDA (Conrad, 2021). → in the process of BDA it refers only to the extraction of relevant data from larger datasets.
Data Cleaning	The cleaning of data aims to identify and remove noise and inconsistencies in data. Necessary steps in this activity are abnormal value detection, filling incomplete data, deduplication and conflict resolution (Ridzuan & Zainon, 2019). These activities are often performed on the basis of statistical evaluation in order to avoid biases in the data. While it often is a tedious process to clean vast amounts of data in order to avoid falsifying insights, it is an integral part for the whole BDA process.

Data Transformation	Data transformation = bringing the datasets towards a single data format. This includes changing the different values towards a single measurement, e.g. transforming different date formats (01.12.2021 vs 12/01/2021). Especially in the case of BDA it includes changing databases stored in different file formats into a format which is compatible with BDA tools in use.
Analytics	Three categories within Analytics: Descriptive Analytics, Predictive Analytics and Prescriptive Analytics Coleman (2016). **Descriptive Analytics:** summarizes and aggregates data in order to make it more easily understandable → visualizations and statistical metrics are primarily used. This form of analysis is used for examining events that happened in the past e.g. a drop in sales due to a competing product. **Predictive Analytics:** focus lays on forecasting of future trends based on historical data. The tools of choice are machine learning, data mining and statistical learning to discover insights from databases. This method also relies on historical data to predict e.g. customer responses to new products. **Prescriptive Analytics:** uses Descriptive or Predictive Analytics and translates it into business decisions → the main methods from optimization theory and operations research are used. It is conducted to determine how business goals can be achieved, e.g. which measures must be taken in order to enlarge the customer base of a certain product.
Interpretation & Evaluation	Interpretation & evaluation = sensemaking and translation of the uncovered insights → the sphere of BDA meets business knowledge. Insights need to be critically assessed towards correlation and causation, confirmation biases and Relevancy (Lebied, 2021). **Example to illustrate the issue of confirmation biases:** the Danish company LEGO. In 2003 the company faced bankruptcy and utilized BD as a countermeasure, with which they determined that Millennials, a core audience for their product, have short attention spans and get easily bored ("Big Data Failure", 2021). As a result they developed larger, more simplistic building blocks instead of their iconic bricks, which further increased their financial decline. They were saved by reconnecting with their customer base and discovered that the mastery of creating their own projects with their toys was more valuable to customers than instant gratification ("Big Data Failure", 2021). This caused them to change their course of action and they now provide an online platform where users can vote for user-submitted projects and products which then will be produced if the interest is high enough. Through that they managed a major turn-around and are now residing in the place of the world's largest toymaker ("Big Data Failure", 2021). LEGO saw their concerns of their toy-bricks not relating to Millennials confirmed through their insight of them having a short attention span. The negative outcome could have been avoided if they would have questioned the correlation and not directly assume causation.

With efficiently integrating the insights uncovered in the BDA process, a company can direct their augmented strategy to obtain the benefits. These benefits will be explained in the following section.

5.2.2 Benefits of BD adoption

Generally speaking, SMEs which increase measurements to use and analyze BD by enhancing technological capabilities consequently elevate innovation, competitiveness, and productivity (Noonpakdee et al, 2018). This is also confirmed by Kalan & Ünalir (2016), which segregate the general benefits of BD adoption into three segments that are mirroring innovation, competition and productivity: make business smarter, increase the revenue, and decrease specific inefficiencies.

Smarter business

In regard to making business smarter, BD can be used to support strategic business decisions with insights uncovered through BDA (Kalan & Ünalir, 2016; Shah et al, 2017; Rajabion, 2018; Iqbal et al, 2018; Ferraris et al, 2019). It can help validate decisions that were previously mostly supported by practical managerial experience (Kalan & Ünalir, 2016; Iqbal et al, 2018; Noonpakdee et al, 2018; Ferraris et al, 2019). And further, BD enables companies to analyze data in real-time, which enables SMEs with a faster and more agile decision-making process (Ferraris et al, 2019). This benefit is achieved through an efficient and thought-through BDA analytics process with respect to the challenges already mentioned. An example for conducting smarter business would be the integration of Netflix's recommendation algorithm, which accounts for 80% of streamed content on the platform (Tozzi, 2021). This resulted in a higher customer retention and has solidified Netflix's position as leading streaming provider. Such a use of BD refers to descriptive analytics, as it is investigated through analyzing the viewing behavior of customers in the past. Additionally, new insights can be uncovered that help understand the underlying dynamics of the business the company operates in (Ferraris et al, 2019). In the same sense, BD can be utilized to gain valuable insights of the foreign market an SME operates in, further driving internationalization efforts (Bertello et al, 2020). Nevertheless, such advances are strongly dependent on conditions that lay outside of the IOSME's influence. For example, BD cannot sufficiently be utilized if the customers of the target market lack behind in digitization efforts e.g. none or less data is generated and thus, not analyzable.

Increase revenue

The enhancement of business decision-making with BD can lead to insights about e.g. new products or untargeted potential customer groups. Integrating these observations into business practice can lead to an increase in revenue. Generally, BD can be performed on social media data, which can support SMEs in understanding customers' preferences and shopping patterns (Rajabion, 2018). By applying descriptive analytics to understand the needs of customers, a SME might be able to provide a differentiated product that satisfies such needs. Such an analysis is even easier for companies that already incorporate purchasing data of their customers. This can be done, for instance, via a clustering algorithm that detects relations in the data of customers and through establishing clusters with similar characteristics (Dittert et al, 2017). By segmenting these customers into separate groups, they can be targeted by using an approach that fits the needs of this group. Here, an example would be Amazon, which generates 30% of their revenue from personalized purchase recommendations (Bertello et al, 2020). Again, such a use of BD falls into the category of descriptive analytics, as it examines past purchasing behavior of other consumers and relates it to consumers interested in an offered article.

Decrease inefficiencies

Lastly, in terms of decreasing specific inefficiencies, BD allows for a better allocation of resources and thus enhances a firm's financial performance (Shah et al, 2017; Iqbal et al, 2018; Ferraris et al, 2019). A use-case example revolves around the problem, that the construction industry accounts 35% of expenses to material waste and work mistakes. Je Dunn, a construction company from the United States, uses BD to provide real-time analytics to their builders, which gives them more accurate projections on the allocation of materials needed for their projects. Thus, less goods are wasted which subsequently improves the financial performance. Further, the vast amount of data further enables the utilization of reliable predictive analytics (Ferraris et al, 2019). UberEats, a subsidiary from the taxi giant which specializes on food delivery uses this to their advantage. By taking into consideration the time it takes to cook individual meals, they are able to accurately predict when a driver needs to pick it up with pinpoint accuracy (Tudor, 2021). This enables the drivers to pick up more than one meal in their delivery trips, reducing the delivery time for their customers and cutting down on inefficient delivery routes. It is also necessary to mention that ongoing efforts towards artificial intelligence and machine learning in combination with BD will result in severe structural change in the future (Mbassegue, et al, 2016). Due to that, computerization will provide companies with the ability to substitute human capital through digital tools (Ulrich et al, 2018).

The earlier a company adapts towards this challenge, the less hard will the impact be when the transition in the industry begins.

In order to determine how these benefits can be reached, the RBV is utilized. Through this theory it can be explained which resources needed to be adapted.

5.2.3 Resource Based View (RBV)

The Resource Based View (RBV) is a theoretical framework that is widely used within international business literature to explain the competitive behavior of organizations. It stems from the initial major works of Barney (1991) in his publication *"Firm Resources and Sustained Competitive Advantage"*, Wernerfelt's (1984) work *"A resource-based view of the firm"* and from Prahalad and Hamel (1997) in *"The core competence of the corporation"* and is still highly applicable today. As the RBV approach has been highly appreciated and widely used within international management, it is a useful framework to examine it in the context of BD. The RBV emphasizes that organizations should consider internal factors to find the sources of competitive advantage and disregard external influences from the competitive environment (Barney, 1991; Wernerfelt, 1984). This is due to external opportunities being more feasible to exploit with existing resources, instead of acquiring new skills for each different opportunity. The theory describes that a firm comprises various resources and by the accumulation of these, competitive advantage can be determined (Wernerfelt, 1984). The RBV is guided by two critical assumptions: resource heterogeneity and resource immobility. While resource heterogeneity means that the composition of resources differs across firms which highlights their uniqueness, resource immobility explains that resources do not move easily across firms and thus are difficult to replicate (Rothaermel, 2013).

Resources include physical, human capital and organizational capital resources (Barney, 1991). In the context of BD, physical capital resources comprise financial resources and software that the firm uses in order to collect, store and analyze data. Human capital involves data scientists and strategists that are capable of capturing information and transform it into beneficial insights and ultimately generate value. Lastly, the organizational capital resources comprise the company's internal structure that enables the company to convert insights into actions (Erevelles, Fukawa & Swayne, 2016).

The financial resources needed for BD adoption are highly dependent on the scope of BD integration. This has implications on data storage, the agility and the speed of the BDA process. The more and the faster a company wants to analyze their analytical problems with BDA, the higher the costs are. These costs are further influenced by the country of operation, which has impacts on e.g. available labor with BD expertise. Additionally, the acquisition of data relates to the size of the customer market. If less information about consumer preferences is available, the higher the price for such data is. To give an example, a package of 50.000 consumer email addresses from the US is equal to 22.000 from France with EUR 35 (Buy Database, 2021). This price is also influenced by the Legal Regulations that apply in the country of operation. The more protected consumer data is, the scarcer it gets and thus, the higher the price for such an acquisition is. Again, depending on the scope of BD integration, the to-be-applied software highly differs in its price. IOSMEs can choose from a wide variety of vendor products with business support solutions or choose to utilize free open-source software solutions.

The suggested human capital used for BD integration in SMEs are three employees with the function of Data Analyst, Business Expert and Leadership. If a company chooses to upskill the necessary capabilities, they can utilize a vast arrangement of free online resources to do so. Nevertheless, they could also employ instructor-led training courses where prices heavily vary between providers and the type of education. Again, a closer look on which analytics problem an IOSME wants to solve can save financial resources in the long run.

The organizational resources are closely linked to the company's corporate culture. Additionally, SMEs are usually less formalized from a hierarchical standpoint when compared to large companies, which gives them the advantage of agility (Bertello et al, 2020). In order to benefit from it, managers must play a key role in promoting the development of BD capabilities throughout the company (Bertello et al, 2020).

Through carefully aligning said resources, the benefits of BD integration can be obtained. These benefits can further lead to a competitive advantage, which will be explained according to the VRIO framework in the following section.

5.2.4 VRIO framework

A competitive advantage is reached when *"it is implementing a value creating strategy not simultaneously being implemented by any current or potential competitors"* while a sustained

competitive advantage further requires that *"these other firms are unable to duplicate the benefits of this strategy"* (Barney, 1991, p.102). To examine if a company's resource can lead to sustained competitive advantage, Barney (1991) has developed the VRIN framework which examines if a resource is valuable, rare, costly to imitate and non-substitutable. The framework has later on been improved to the well-known VRIO framework, adding the question of whether a company is organized to capture the value of the resource (Rothaermel, 2013). The steps within the VRIO framework that need to be acknowledged within a RBV are illustrated in Figure 10.

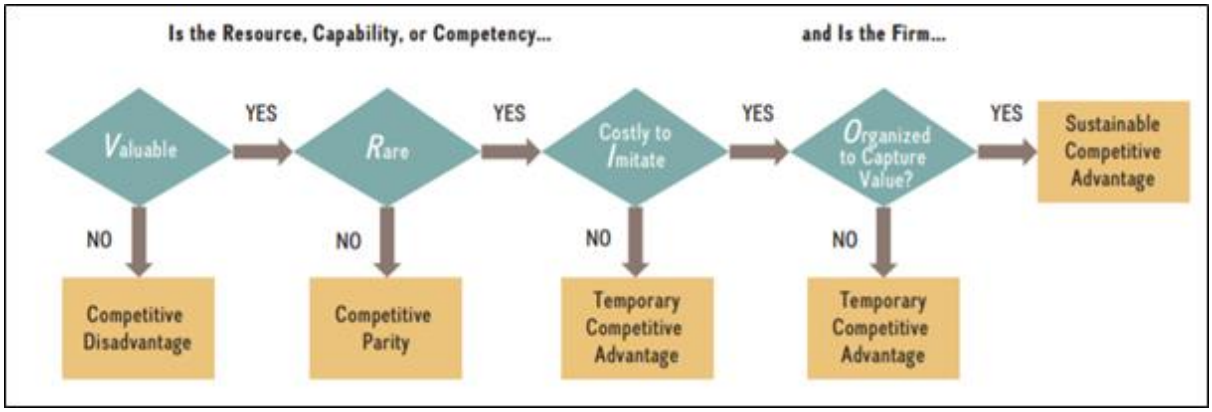


Figure 10: VRIO framework in the RBV

Source: (Rothaermel, 2013)

The first step relates to the question of value. A resource is considered valuable if the firm can consequently increase value for its customers by either increased differentiation or decreased costs. If the resource is not considered valuable it leads to competitive disadvantage. Secondly, a resource is considered rare if only a few firms possess it. When rarity is not given, competitive parity is reached. If the resource is considered valuable and rare, at least temporary competitive advantage can be achieved. However, thirdly, the resource must also be costly to imitate or substitute for a competitor if the aim is to reach sustained competitive advantage. Lastly, the company needs to be organized to capture the value from the resource, meaning it needs an effective organizational structure. Only if all these four criteria are met, the resource can lead to sustained competitive advantage (Rothaermel, 2013).

5.2.5 Competitive advantage in the context of BD

In order to achieve a sustainable competitive advantage, a complex mix of different financial, human, physical, and organizational resources need to be employed in addition to BD (Ferraris

et al, 2019). Through this, superior and difficult-to-imitate BDA capabilities can be created which in turn enhances firm performance (Ferraris et al, 2019). Then, IOSMEs are able to leverage the full potential of BD and consequently gain a competitive advantage in internationalization which will positively impact foreign sales (Bertello et al, 2020). However, BD governance itself does not lead to a competitive advantage (Ferraris et al, 2019).

The faster the decision-making for an IOSME can be achieved, the more BD becomes a competitive advantage. And as more data can be stored, it allows for a deeper BDA through enriching the analysis with more historical data (Kalan & Ünalir, 2016). Such an analysis would lead to more stable and thus more trustworthy results. An example for a company achieving a sustainable competitive advantage is Spotify. Spotify provides their customers with better playlist recommendations to its customers. Through the millions of their users already listening, they have the necessary data to utilize data analytics algorithms to gain insights into user preferences. They use natural language processing to analyze the language, lyrics, and contents of a song (Valcheva, 2021). Natural language processing is used to read, decipher, and understand textual data to make sense of it. Possible applications are sentiment analysis on Twitter Tweets to quickly determine the general mood towards a certain topic. If, for example, generated tweets towards a certain topic (hashtag) contain a lot of swear words, then this process would show a correlation towards a negative sentiment. Next, they utilize raw audio analysis to describe the mood of a song, which can be e.g. upbeat, instrumental, chill, etc. (Valcheva, 2021). Through that they attain a user profile which stores the preferences of a user. Lastly, they use collaborative filtering to compare new songs to the user profile and his listening habits to decide if it should be recommended (Valcheva, 2021). Thus, they can offer a personalized playlist with music recommendations based on their listening history, which gives them a sustainable competitive advantage that relates to a higher user retention.

While the mentioned benefits and potential to reach competitive advantage through the use of BD are undoubtedly compelling for IOSMEs, the integration is subject to different challenges which will be explained in the following section.

5.3 Challenges of BD implementation

This part aims to shed light on the general barriers for an integration, followed by obstacles that specifically IOSMEs could encounter. These insights will lay the basis to understand which capabilities an IOSME needs to change to successfully adopt BD. After reviewing the relevant

set of articles, the barriers for BD adoption in IOSMEs can be categorized into internal and external challenges.

5.3.1 External challenges

The external challenges are considered as Data Quality, Legal Regulations and Technological Opacity. These challenges will be further explained in detail with respect to their affiliation towards IOSMEs and their possible solutions.

Data Quality

A successful implementation of BD is highly dependent on the composition of data. While the data used in businesses is generally in a structured format, the data structure of external data sources is governed by factors outside of the IOSMEs influence. With the various forms of BD related to the Variety of the 5V's, it must be ensured that the gathered data stemming from multiple sources can be processed, examined and analyzed. A factor which takes this in regard is Data Quality (Rajabion, 2018; Ferraris et al, 2019). In general, Data Quality ensures that the data can be used as a supportive function for business decision-making (Kalan & Ünalir, 2016). As a multi-dimensional concept, Data Quality not only depends on a sufficient technical implementation such as a complementing IT infrastructure, but also relates to multiple standards the data itself needs to fulfill. According to the definition of Data Quality by Cai & Zhu (2015), these standards are Availability, Usability, Reliability, Relevance and Presentation Quality, visualized in Figure 11. The first four dimensions are indispensable features to provide Data Quality, with the fifth improving the understanding of the data (Cai & Zhu, 2015). These concepts ensure that the challenge of Data Quality is met will be explained in the following parts.

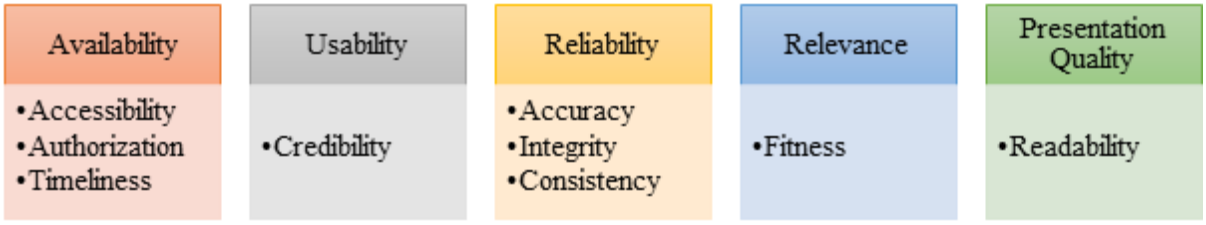


Figure 11: Data Quality Dimensions
own compilation, based on (Cai & Zhu, 2015)

Concerning Availability, an IOSME has to make sure to make the data accessible for authorized employees in order for it to be analyzed (Kalan & Ünalir, 2016). The Authorization is the basis for the governance of data. Tracking which employees have access to which data, is a main aspect of data security. To give an example, company data including the wages of employees should only be accessible for departments that have to process such data. The Velocity of BD, mentioned in the 5V's further adds to a shorter lifetime and thus, a smaller window of opportunity to profit from it (Kalan & Ünalir, 2016). Thus, the Timeliness of data relies on regular updates in the data and if the time interval between data collection and data processing meets the companies requirements (Cai & Zhu, 2015). This is even more important if a company wants to conduct real-time analysis, where the data has to be immediately available for further processing.

Usability is concerned with whether the data is useful and meets the needs of users (Cai & Zhu, 2015). Usable data for a specific problem might be scarce, depending on e.g. the industry or the countries an IOSME operates in. The less data is available, the more important it is to check on the credibility of the data. This credibility can be further enhanced through field matter experts that check the correctness of the data content (Cai & Zhu, 2015). Through that, a company can ensure that the used source of data ensures a certain level of Data Quality in their set which in turn refines the acquisition process of data.

Reliability relates to the trustworthiness of data (Cai & Zhu, 2015). First, it needs to be ensured that the provided data is accurate and reflects the true state of the source information (Cai & Zhu, 2015). This can be examined by testing the dataset towards unrealistic or missing values. A main requirement for extracting insights from BD is to minimize the complexity of its nature (Iqbal et al, 2018). With BD stemming from social media, newspapers, and other sources, it can lead to data integration complexities when sourcing them together (Ferraris et al, 2019). Such errors of integration can damage the integrity in the dataset, leading to flawed results when not checked. Further, results from BD have to be proven towards consistency. A conducted analytical process has to deliver the same results every time it is conducted (Cai & Zhu, 2015). Thus, the accuracy revolves around the data reflecting the true stage of the source information over time.

The Relevance is concerned with the correlation between data content and the analyzed BD problem. The Fitness of data ensures that the data used for a specific analytics problem either

matches the theme of this problem or expounds at least one aspect of the theme (Cai & Zhu, 2015). With BD centering around multidimensional data sources, analysts might be able to uncover a relationship between an increase in revenue and the weather forecast. But here, the old saying ‘correlation does not imply causation’ applies, meaning that two causes do not singlehandedly justify an outcome.

Presentation Quality is equivalent to the readability of results from the analytics process. Here, it means that the insights which were uncovered are explained in known or well-defined terms (Cai & Zhu, 2015). As interpretation and evaluation plays a key role in the analytics process, this point highlights the ability to translate the insights uncovered through complex methods into an understandable format.

Legal Regulations

Depending on which type of data is used, Legal Regulations pose a barrier for BD adoption as well (Rajabion, 2018; Iqbal et al, 2018). Coleman et al (2016) segregates the aggregation of BD into three main sources: human sourced data (unstructured data, records of human experience e.g. social network data), process mediated data (structured data, coming from traditional business systems e.g. commercial transactions) and machine-generated data (metadata, measures from the physical world e.g. GPS-logs). This plays into the Variety of the 5V’s, which enriches possible insights due to multidimensional aspects and thus, strengthens the insights uncovered in the BD process. While machine generated data can be even made public if the company wishes to do so, human sourced data is subject to vast and complex Legal Regulations which depend on the country of operation. A famous example is the Cambridge Analytica scandal from March 2018, where they acquired and used personal data from up to 87 million Facebook users without their consent in order to sway public opinions by psychographic ad-targeting ("Cambridge Analytica", 2021). This breach of Legal Regulations and ethics resulted in the closure of the firm in 2018 ("Cambridge Analytica Aftermath", 2021). Nevertheless, BD can also lead to significant positive effects in various industries such as transportation, healthcare, and energy, but face the problem of being highly regulated for security purposes (Rajabion, 2018). Such regulations are not easily understandable for judicially untrained persons (Coleman et al, 2016). This challenge is further increased if the SME who wants to adopt BD has little or no in-house legal expertise (Iqbal et al, 2018).

These Legal Regulations add to the slow adaptation of BD in SMEs, as the OECD (2019) argues for a higher dependency on business ecosystems and policy environment when compared to larger businesses. And further, larger business entities, such as multinational enterprises (MNEs), generally have more financial resources available for integrating BD into their business practices, which is a crucial point for an adoption. Additionally, a MNE generally has a bigger customer base and might reap the benefits of BD fairly easier than a smaller company could. This circumstance is paired with the general slow adaptation of BD in IOSMEs, which leads to the 'bigger players' dominating whole industries in BD utilization. But as SMEs and IOSMEs play a vital role for the economy overall, insights to help them overcome integration challenges of BD are of growing importance. However, SMEs that engaged in new opportunities provided through BD have experienced a double- or triple-digit growth in many countries, further proving the possibility for SMEs to greatly benefit from an early adoption (OECD, 2019).

The examined literature suggests that SMEs would need to outsource this resource by assigning third-party legal support in order to abide by the existing rules. However, this is cost expensive and goes against the limited financial resources of SMEs (Coleman et al, 2016; Iqbal et al, 2018). Another option would be to utilize cloud computing services, which provide service level agreements that ensure data security.

Technological Opacity

A challenge mentioned especially for SMEs is Technological Opacity (Rajabion, 2018; Dittert et al, 2017; Kalan & Ünalir, 2016; Iqbal et al, 2018; Coleman et al, 2016). It considers the difficulty for SMEs to understand the usage and benefits of new technologies for their companies (Coleman et al, 2016). This leads to them remaining doubtful of making investments into technologies they do not know and/or understand (Coleman et al, 2016; Rajabion, 2018). But this challenge not only relates to the concept of BD, as is the case for this thesis, but also includes the problem of choosing the right tools. The rapid growth of BDA in the software market further increases the difficulty to get an overview over possible options. Additionally, available evaluation platforms for such tools are strongly vendor biased (Coleman et al, 2016). This poses a challenge for IOSMEs as a case-sensitive and independent evaluation becomes problematic. Again, this influences the general adaptation rate of IOSMEs towards BD.

Dam et al. (2019) suggest that in the early stages of BD adoption the relevant scale and scope of data and the sophistication of BDA techniques should be in focus. Although there is the possibility of engaging with software vendors which would mean that this challenge can be overcome through support from outside of the company, the initial step is to examine the needs of the BD process a IOSME wants to integrate. As mentioned before, by determining which analytical problem an SME wants to address with the usage of BD beforehand, the choice in utilizable tools and processes become clearer (Kalan & Ünalir, 2016). Through that, IOSMEs can carefully examine the available options for a successful integration of BD with respect to the aforementioned features of the BDA software market.

5.3.2 Internal challenges

The internal challenges consist of choice in IT infrastructure, the retrogressive Company Culture and BD expertise. In the following sections it will be explained what these challenges are in detail, why they pose barriers for IOSMEs specifically and how they can be overcome.

IT infrastructure

The primary requirement for a successful implementation of BD into business practice is an adequate IT infrastructure which is able to handle BDA processes (Ferraris et al, 2019; Kalan & Ünalir, 2016; Iqbal et al, 2018; Coleman et al, 2016). In order to handle the vast amount of data processed, the existing IT infrastructure needs to be reconfigured. Generally, the technology for storing and processing data is involved with high up-front and operational costs (Kalan & Ünalir, 2016). Further, the limited financial resources of SMEs make them cautious about new investments beyond their business scope (Coleman et al, 2016). This is coherent with the literature review, as the most prominent factor mentioned for keeping SMEs from utilizing BD were costs (Rajabion, 2018; Dittert et al, 2017; Kalan & Ünalir, 2016; Iqbal et al 2018; Bertello et al, 2020). It also has to be mentioned that the more experimental an SME wants to be in the possible applications of BD, the more flexible the IT infrastructure has to be, which in turn again elevates costs even more (Wang & Wang, 2020). These additional costs can be fatal for SMEs with very limited budgets and expertise (Coleman et al, 2016). Especially for IOSMEs, scalability is of high importance, as new subsidiaries might be incorporated into a company-wide BD concept. Being able to scale an existing data-driven system by integrating other sub-companies into the network and profiting from the expansion should be focused on when adopting BD. Thus, it has to be made sure that the IT infrastructure can be extended on demand. Wang & Wang (2020), further state the problem of overuse in IT infrastructure for

SMEs. This is a consequence of an overestimated vision of their BD utilization. In this case, a point will be reached where the cost of storing and processing data will become too expensive for an SME with limited financial resources (Kalan & Ünalir, 2016). Further, Data Security is an often-mentioned point in itself throughout the examined articles and must be considered when setting up an IT Infrastructure (Rajabion, 2018; Ferraris et al, 2019; Dittert et al, 2017; Iqbal et al, 2018; Coleman et al, 2016). It is easier achievable for larger corporations but more cost-involved for smaller enterprises. SMEs specifically overemphasize on data security in comparison to larger companies, in order to protect organizational reputation (Rajabion, 2018; Iqbal et al, 2018). The need for an established efficient data security for SMEs is further implied by Coleman et al (2016), as they state that 40% of all cyber-attacks were directed at SMEs. To name an example, a major security risk for many SMEs are outdated database management systems whose developers have stopped the support (Iqbal et al, 2018). Under these circumstances, SMEs face significant threats of cyber-attacks e.g. data breach or intrusion (Iqbal et al, 2018). Consequently, it needs to be ensured that the utilized IT infrastructure is protected against unauthorized access.

Wang & Wang (2020) refer to a practical strategy to determine the scope as *"think big but start small"* (Wang & Wang, 2020; p. 2). Following this advice, an IOSME would want to consider using cloud computing services when setting up an IT Infrastructure. Wang & Wang (2020) further state that the strategic utilization of BD should align with the business strategy and embrace a long-term plan. Thus, the level of investment in IT infrastructure is highly dependent on the initial scope of utilization of BD. A company seeking to adopt BD into their business processes needs to determine which analytical questions it wants to answer with it beforehand. The scope is determined by a multitude of factors, such as the size of data stored, the agility of the analytics process, maintenance tasks, and so on. Additionally, next to the initial costs, the future tuning and performance costs when scaling the existing IT infrastructure also need to be considered. Even more so when the ongoing process of BD implementation is slow in reaping benefits. Thus, before setting up an adequate IT infrastructure a company has to pre-determine the scope of its BD utilization and options to scale the existing infrastructure when necessary.

Company Culture

SMEs tend to specifically construct their business model towards a specific market opportunity or inhabit resources that provide them with a competitive advantage in their respective markets. Such SMEs are perceived to be domain specialists in their respective industry which leads to

them overlooking opportunities they would have at their disposal (Coleman et al, 2016). This risk increases in combination with less developed general management functions due to the smaller hierarchy present in SMEs (Iqbal et al, 2018). The result is a disregard for new concepts such as BD (Dittert et al, 2017). While this can be seen as closely connected to the challenge of Technological Opacity, its main difference is that Technological Opacity centers around a lack of knowledge while the Company Culture often is related to ignorance. This is confirmed by Noonpakdee et al (2018), who observed that if the top management in an SME is part of the baby boomer generation, they often do not believe that BD could help them attain competitive advantages. This lack of awareness can lead to a retrogressive Company Culture, which actively sustains new technological implementations and sticks to 'business as usual'. This can also be due to BD's ability to diminish the value of managerial experience through supporting business decision-making based on facts rather than intuition, which in turn leads to opposition against this new technology. Such an internally created barrier is a main factor of why SMEs perceive BD more as a hype rather than a transformative technology (Rajabion, 2018). Some SMEs perceive BD as not worth implementing due to their own smallness (Noonpakdee et al, 2018). The potential benefits are disregarded when compared against the necessary organizational implementations. Another influence is the doubt of SMEs if the data used in the company even falls into the BD category at all (Noonpakdee et al, 2018). And lastly, a common mistake when adopting BD practices is to treat BD as a delimited over time project rather than a continuous exploration device (Shah et al, 2017).

To fully obtain the benefits of BD it is necessary for managers to align existing organizational culture and capabilities across the whole organization (Ferraris et al, 2019). This can be done through the employment of a BD agent that promotes BD usage by experimenting with different capabilities, asks provocative questions and who guides the SME and its employees in the transformational process (Coleman et al, 2016). Further, the firms' attitude towards managing and integrating new and old knowledge can improve the positive effect BD has on the company's performance (Ferraris et al, 2019). It is also mentioned that the key challenge of using BD is to make the usage of it understandable and trustworthy for all employees (Ferraris et al, 2019). Only if employees at all levels are understanding and including data in their decision-making process the benefits of BD could be materialized (Ferraris et al, 2019).

BD expertise

The lack of BD expertise is mentioned as another primary reason for slow BD adoption in SMEs (Rajabion, 2018; Ferraris et al, 2019; Dittert et al, 2017; Iqbal et al, 2018, Coleman et al, 2016).

In order to elevate the current handling of data to fit the nature of BD and its implications from the 5V's, the available capabilities of the IOSME may be insufficient and thus might need to be reconfigured. Naturally, the application of BD tools requires technical skills. Reitter (2021) suggests a team of at least three employees for a smaller company structure in order to provide the technical and the managerial expertise needed for DB implementation. Their roles are Data Analyst, Business Expert and Leadership. The role of the Data Analyst revolves around the technical expertise needed when implementing BD. The Data Analyst provides implementation services for BDA processes, guidance in the use of BDA techniques and system support. This role is of paramount importance in the initial phase of BD implementation. The Business Expert incorporates extended knowledge of the business's operating procedure and systems. This role is concerned with aligning the business requirements to the used BDA processes. Further, it can also be seen as a field matter expert checking on the Usability aspect in the aforementioned Data Quality. The Leadership position understands the importance of data as a strategic business tool. This position takes an active role in prioritizing BDA tasks and exemplifies the consistency and discipline needed for a strategic application of BD. The Leadership position relates to the BD agent promoting BD usage throughout the company in order to improve the Company Culture towards a data-driven business. There is also the possibility of employing only one person that accomplishes these roles, although it would heavily affect the balance of work distribution and diverse insight (Reitter, 2021).

While larger companies are able to employ a larger number of human capital and distribute multiple functions among them, SMEs need cross-functional experts which have expertise in multiple domains (Iqbal et al, 2018). This is due to the complexity of setting up, maintaining and using a functional IT infrastructure in a smaller team constellation, with tasks that range from database administration, hardware knowledge, programming expertise and analytical skills (Coleman et al, 2016). Nevertheless, such 'all-rounders' are even harder to find in the limited labor market (Iqbal et al, 2018). To give an example, the United States faces a current shortage of up to 190.000 data analysts and another 1.5 million managers and analysts with necessary skills to abstract decisions based on BD (Coleman et al, 2016). This shortage is further intensified as the lack of available BD specialists increases the salaries to levels which are not feasible for most SMEs (Coleman et al, 2016). To exemplify the necessary financial resources to obtain a team consisting of three experts, online wage references according to the position of a Data Analyst, Business Process Consultant (Business Expert), and Senior Management Consultant (Leadership) were examined. The expected wages for employing such

a team, in the example case of Denmark as a targeted point of operation, round up to 197.000 EUR annually ("PayScale", 2021). Larger companies often circumvent that problem by taking over smaller analytical firms (Coleman et al, 2016). However, a takeover might not be possible to the same degree for IOSMEs due to their limited financial resources.

The European Commission (2019) states that SMEs require necessary support to train their employees in digital advances if they want to thrive in a data-driven economy and enlarge their profit with the usage of BD. But such upskilling is only one solution to the barrier in lack of BD expertise, next to developing a own BD department and outsourcing. When upskilling, IOSMEs could consider utilizing open-source software tools which can handle BD. Such tools are often complemented with a vast online community which offer free open online courses and learning material (Coleman et al, 2016). Through this, internal capabilities can be improved without investing vast financial resources. When developing an in-house BD department, IOSMEs could consider training the current IT staff in BD techniques (Coleman et al, 2016). With respect to the limited financial resources of IOSMEs, Wang & Wang's (2020) quote applies to such an effort with 'think big but start small'. The last option is to rely on external expertise, which is costly (Dittert et al, 2017).

With the difficulty of obtaining BD expertise externally from the already exhausted labor market or internally with the need for 'all-rounder' specialists, the challenge of BD expertise can be perceived as being an internal as well as an external challenge for IOSMEs. Subsequently, a IOSME could choose if it wants to overcome this challenge by building up existing resources or by utilizing external resources, if available.

5.3.3 Additional barriers for IOSMEs

A questionnaire from Bertello et al (2020) answered by 103 SMEs resulted in the finding that the relationship between governance of BD infrastructure and the internationalization performance is not significant. However, it also resulted in the insight that BDA capabilities directly affect SMEs international growth (Bertello et al, 2020). This means that BD governance itself is not enough to enhance internationalization in an IOSME, as it needs to be combined with BDA capabilities in order to have a positive impact on an IOSME's growth. Thus, the following section will explain BDA in order to provide an understanding of how these benefits can be achieved.

Traditionally, SMEs are focused on conducting business in their home markets and orient their strategy accordingly to only one market. In the case of IOSMEs, their strategy cannot be applied to all markets in the same strength. Therefore, other aspects need to be considered as well. Here, the Cultural Context of IOSMEs doing cross-country business across different contexts influences the strategy a IOSME wants to utilize.

Cross-cultural context on BD utilization

IOSMEs conducting business in different countries have to deal with a variety of national and regional cultures. This has an influence on customer behavior whether a company is aware of it or not. It also can impede the expansion into foreign markets and can lead to failure in their venture when misjudging the social, cultural and political environment ("Big Data Failure", 2021). Thus, an IOSME has to weigh its impact on their organization and BDA process. Such regard for cultural influence plays a lesser role when dealing with non-human data but is vital when dealing with consumer-related decision-making. As mentioned before in the challenge of Data Quality, it is necessary to fit the theme of analysis to the utilized data. Apart from technological and logical aspects in the analysis, data itself is generated within a sphere which itself is influenced by the underlying culture and morals of individuals. This was the case for a western company who wanted to research business perspectives during the first week of May in Japan. They oversaw that this was the 'Golden week', a collection of holidays which in turn led them to not being able to access the right business representatives within their time-frame (Tian & Tobar, 2004).

Generally, the research of how cross-cultural work influences technological disciplines is still in its infancy (Park et al, 2010). Prescott & Gibbons (1993) identified five issues that can disrupt efforts in gathering data in a global setting. Firstly, the types, timeliness, accuracy and motives for data collection varies between different cultures. Thus, it is necessary to identify which Data Sources produce the most timely, accurate and reliable data. Secondly, the ethical standards of data collection and how such individuals perceive data collecting is different between cultures. While western economies tend to prevent IT companies from gathering information without the users consent, which is inherently different in Asian countries. Especially in China, where the state gathers information about their citizens with various techniques. Thirdly, the technologies for production, storage, movement, analysis and timing of data are significantly different across countries. The usage of information technologies increases globally, but the grade of IT adoption differs heavily between countries and cultures. Fourthly, language barriers impede the

collection and analysis of information. Fifthly, culture-specific idiosyncrasies must be addressed.

Additionally, an IOSME could refer to competitors' attempts in adapting towards a foreign market to gain more information (Prescott & Gibbons, 1993). Here, potential blind spots in the competitors and one's own strategy can be uncovered and utilized for a potential advantage. Park et al (2010) further mentions the possibility of using a Cultural Simulation Modeler. It tool was initially constructed as a terrorism assessment tool and has been made commercially available as a business intelligence software (Park et al, 2010). It creates a cultural construct by filtering data sources with the help of subject matter experts to assess contextual information towards a specific issue. Through that, cultural awareness can be enhanced. While such a tool is currently not feasible for IOSMEs, it has to be noted that advances in this field can lead to tools that might be implementable without the utilization of vast financial resources.

Internationalization

Internationalization plays an increasingly important role in the economic development for such companies. With the EU adopting over 280 policy measures that support the internationalization of SMEs since 2011, the opportunity to engage in internationalization gets increasingly easier to access. A further statement of nearly all EU member states having increased their internationalization efforts since 2008 shows the vast improvement made in this area (European Commission, 2019). Thus, the European Commission has realized the need of SMEs to adopt BD practices and urges them to investigate data-handling tools and methods (Coleman et al, 2016). IOSMEs have a higher potential for generating new opportunities and benefits through BD when compared to SMEs. This is due to more information being available for an analysis and thus, there is a greater leverage for benefiting from it. An example for application in IOSMEs is e.g. by enlarging their market for selling products or expanding their customer base, they have a higher incentive for an integration (Swoboda & Foscht, 2014).

When internationalizing, there are possible uncertainties from various factors such as information asymmetry, geographical distance, and the difficulty both of enforcing contracts across borders and of ascertaining the capabilities of a foreign distributor (Bertello et al, 2020). Additionally, small enterprises are more challenged by committing resources to foreign markets, which is why they should gradually adjust their level of commitment, based on their

specific advantages on the local market (Dam et al, 2019). The higher the commitment, the more superior is the internationalization performance (Dam et al, 2019).

One should also mention that the market competition has expanded and transcends geographic boundaries (Mbassegue, et al, 2016). Thus, ignoring the possibility of BD utilization on making better decisions will further weaken SMEs (Mbassegue, et al, 2016). Here, BDA can make the difference by exploring and scanning the market and assessing risks (Bertello et al, 2020). It enables IOSMEs to make informed and effective decisions on choosing which foreign markets are worth investing in, based on related opportunities and threats arising from external trends and changes (Bertello et al, 2020). As IOSMEs face the liability of smallness and liability of foreignness during their internationalization process, BD can help address these problems (Bertello et al, 2020). BD is able to stimulate international growth in IOSMEs by lowering uncertainty of foreign expansion (Bertello, et al, 2020). This is especially vital in an early stage of internationalization for IOSMEs (Bertello et al, 2020). Thus, BD enables enterprises to find better strategies and thus enhances their performance in internationalization (Dam et al, 2019). In order to use BD for an increase in internationalization performance, enterprises need to build capacities in knowledge management (Dam et al, 2019). As mentioned before, knowledge management perceives knowledge gained from BDA as a source of competitive advantage (Dam et al, 2019). By acquiring, assimilating and sharing knowledge gained from BDA, the Value from BD can be leveraged (Dam et al, 2019). It is further necessary to mention that knowledge management plays a critical role in BDA application in strategic, tactical and operational areas, as the influence of knowledge and the impact of BDA are correlated (Wang & Wang, 2020). In combination with the previously mentioned potential improvements in other areas, SMEs would benefit from adopting a data-driven perspective in organizational culture (Bertello et al, 2020).

To conclude, the internal and external challenges state the requirements for a successful BD implementation in an IOSME. Through carefully investigating the needs of the analytical problem an IOSME wants to solve and by orienting the solutions accordingly, the stated benefits can be achieved. These relationships in regard to BD implementation are summarized in Figure 12.

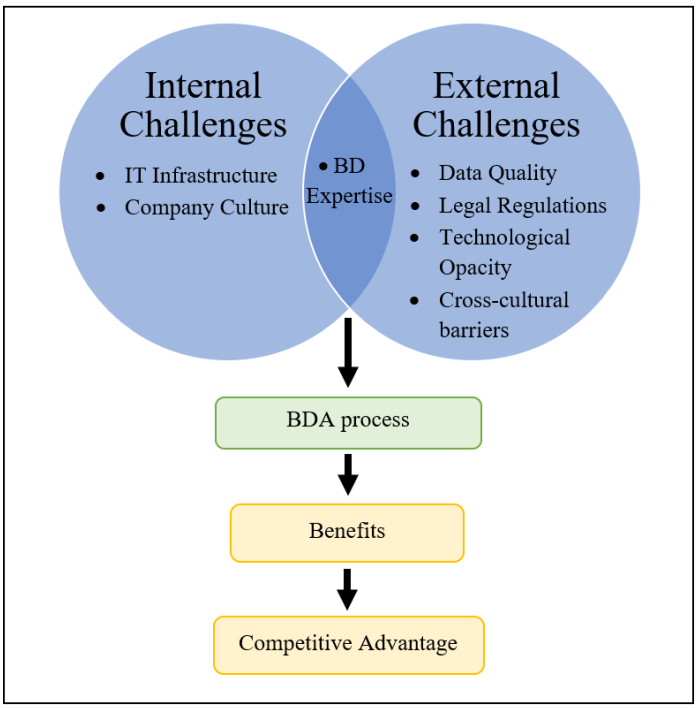


Figure 12: Summary of BD implementation

own compilation

5.4 Options for BD implementation in IOSMEs

Here, it will be examined which possible internal and external options for BD adaptation in IOSMEs according to the conducted literature review exist. Additionally, this section will focus on possible advantages and disadvantages of these options in order to provide a clearer picture for the reasons of integration and where an application is beneficiary. The answer will be divided into internal and external solutions for integrating BD in IOSMEs. Following, the Resource Dependency Theory will be explained in order to provide a theoretical foundation for the utilization of external services and its implications on competitive advantage.

5.4.1 Internal Solutions

The internal solutions prominently mentioned in the literature review were open-source software and cloud computing. Together with external solutions, they will be thoroughly examined according to their implications of utilization, advantages and disadvantages in order to investigate their role in BD implementation in IOSMEs.

Open-source Software

According to the challenge of Technological Opacity, a specification of the analytical problem can support IOSMEs in their investment decisions for tools and skills. Hence, Rajabon (2018)

suggests a thorough examination of the available tools on the market to help identify the best technical solution for an IOSME. Nevertheless, companies that provide comparison and evaluation of BDA software for SMEs tend to favor certain BDA software developer companies (Iqbal et al, 2018). Comparing BDA products on functionality alone is challenging, as many tools share the same features and capabilities (Coleman et al, 2016). A company needs to be able to correlate ease of use and algorithmic sophistication to the organization's capability and level of maturity in analytics (Coleman et al, 2016). The choice in tools for a BD solution must be able to offer deep analytic, high agility and a good potential for scalability in combination with low latency and costs (Kalan & Ünalir, 2016). Such tools enable IOSMEs with the ability to shift between examined problems and to stay agile in their application of BDA.

But this decision does not have to be oriented towards costly solutions, as a significant part of articles mentioned the utilization of free open-source software for IOSMEs (Coleman et al, 2016; Kalan & Ünalir, 2016; Dittert et al, 2017; Noonpakdee et al, 2018; Dam et al, 2019; Wang & Wang, 2020). Open-source software relates to the publicly available source code of a software with which a software can be inspected, modified and enhanced ("Open source", 2021). A range of well-established open-source solutions exist for each step in the analytics process, which can be seen in Figure 4 (Coleman et al, 2016). Some of the available open software solutions are as powerful as competitive commercial products for data collection and processing (Wang & Wang, 2020). And further, as these open-source software tools often gather a vast supporting online community, SMEs can benefit from free available training material and advice (Wang & Wang, 2020). These two factors directly play into the limited financial resources of IOSMEs and the challenge of BD expertise.

Hadoop, an open-source software tool which spawned an entire industry of related services and products, was the most remarked BDA tool in the examined literature (Coleman et al, 2016; Kalan & Ünalir, 2016; Shah et al, 2017; Noonpakdee et al, 2018; Rajabion, 2018; Ferraris et al, 2019; Wang & Wang, 2020). The Hadoop open-source framework enables processing, storing and analyzing vast amounts of distributed data in any format (Ferraris et al, 2019). Hadoop is based on a distributed file system, which breaks larger sized files into smaller blocks and distributes it across the connected machines while processing them parallelly (Dikshantmalidev, 2021). This gives Hadoop an edge against other software solutions in terms of flexibility and speed in the analytics process. Additionally, Hadoop can be run on commodity hardware (Dikshantmalidev, 2021). Such off-the-shelf hardware translates to less expenses as

it is not necessary to buy specialized hardware in order to run the software. This further enables an IOSME to inexpensively scale the existing system when necessary. Lastly, Hadoop replicates the data it handles by default across its utilized machines, leading to a high fault tolerance in the case of one machine crashing (Dikshantmalidev, 2021). Thus, a faulty machine can be replaced and troubleshooted without hindering the advance of the analytics process.

However, Hadoop also has its downsides. While it is efficiently storing and processing BD, it faces problems in processing vast amounts of smaller size data which results in performance and processing issues (Dikshantmalidev, 2021). Depending on the data an IOSME wants to utilize this can be a knock-out criterion for utilizing Hadoop. This software also has critical issues in terms of Data Security. As Hadoop is written in the Java programming language, the most commonly used programming language, it faces an increased threat of possible cyber-attacks (Dikshantmalidev, 2021). Additionally, the default security feature in Hadoop does not utilize storage and network encryption, raising the thread in Data Security even more (Dikshantmalidev, 2021). Lastly, Hadoop only supports batch processing, meaning new data needs to be processed before being available for analysis (Dikshantmalidev, 2021).

To give a practicable example, Staple, a US supply chain store, utilizes Hadoop to process 10 million data transactions every week to forecast sales across their 1.100 retail outlets through which they were able to reduce their market promotion costs by 25% (Sistla, 2021). This shows the applicability of Hadoop towards extensive data environments. However, an IOSME would have to examine the data which will be applied in the analytics process beforehand in order to determine the applicability of Hadoop.

Cloud Computing

As part of the integration of BD management in SMEs, many articles have referred to utilizing cloud computing (Coleman et al, 2016; Kalan & Ünalir, 2016; Dittert et al, 2017; Rajabion, 2018; Iqbal et al, 2018; Noonpakdee et al, 2018; Dam et al, 2020; Wang & Wang, 2020; Schuilling, 2020). Cloud computing is defined as a computing model which ensures a real-time connection to a shared system available to many devices (Rajabion, 2018). As such, it substitutes the necessary IT infrastructure through which BDA can be performed. By using cloud computing services, an IOSME is able to use servers that are deployed and maintained by third parties (Coleman et al, 2016). This translates to reduced costs of software ownership in licensing and technical maintenance (Coleman et al, 2016). As the challenge of IT

infrastructure correlates to high investment costs, this solution complements the limited financial resources of IOSMEs as well.

Its benefits for IOSMEs can be summarized into three categories: Efficiency, Computational advantages and Privacy & Security (Rajabion, 2018). The Efficiency relates to the IOSMEs performance in operations and data management, as Cloud Computing offers a collective database which can be easily shared across with multiple devices and users (Rajabion, 2018). This benefit stands out for IOSMEs as it offers an easy integration of additional subsidiaries into the IT infrastructure. This is important, as a pooled dataset enriches and strengthens the insights uncovered in the analytics process. On Computational advantages, the most important factor is the possibility to access more extensive environments with increased processing power which potentially gives an IOSME enough leverage to compete with bigger organizations in the industry (Coleman et al, 2016). Such extensive environments are concerned with the ability to easily enhance the currently used computational power or data storage capacity. Being able to enhance the processing power by simply extending a current service plan allows IOSMEs to fluently adapt its resources when needed. This higher available processing power further allows for more rapid analytical capabilities (Kalan & Ünalir, 2016). Again, this lowers the reaction time towards uncovered insights which positively affects the potential competitive advantage. The Privacy & Security aspect relates to automatic software upgrades, license management, high reliability and protection through customizable security protocols (Coleman et al, 2016). To ensure Data Security, service level agreements are utilized to clarify the responsibility between cloud service providers and cloud service customers (Kalan & Ünalir, 2016). These service level agreements often are subject to varying applicable legal requirements, depending on e.g. if personal data is hosted in the cloud service (Kalan & Ünalir, 2016). This allows for agile changes in the analytical process and enables IOSMEs to experiment with their scope of BDA. Additionally, such service level agreements ensure business continuity through disaster recovery (Kalan & Ünalir, 2016). As IOSMEs face a higher threat of cyber-attacks, it has to be noted that most cloud services provide an affordable data backup and replication service, which makes this solution even more attractive for them. Ferraris et al (2019) states that due to lower costs in regard to storage, memory, processing, bandwidth and so on, SMEs that are not born digital will become increasingly able to engage in BD. This might become an even more important factor in the future as computational power rises and the costs drop steadily.

Although cloud computing seems like a perfect solution for IOSMEs, a survey showed that 35% of SMEs surveyed use the cloud for data storage while 42% do not use the cloud at all (Kalan & Ünalir, 2016). A further 27% stated that they do not understand the concept of cloud computing well or at all (Kalan & Ünalir, 2016). These numbers can be related towards the challenge of Technological Opacity. Additionally, SMEs often defer the option of cloud computing due to the data storage residing outside their on-premise data centers, which increase risks of accessibility, confidentiality, and privacy (Kalan & Ünalir, 2016). Thus, data would be delocalized. And further, as more data needs to be transferred over the network it slows BDA performance, whereas data locality would improve it (Kalan & Ünalir, 2016). Nevertheless, technological advancements in computing power have constantly improved, also positively affecting bandwidth and reducing data transfer time over online networks (Mbassegue et al, 2016). Additionally, for an IOSME with at least one subsidiary, an online service which pools the available data together is more feasible. Sourcing the available company data into a single data warehouse is also of high importance when utilizing BD. If a company stores its data in separate entities, the outcome of BDA processes would differ depending on which data was utilized. Additionally, data centers based on a cloud architecture are distributed over a large geographical area (Coleman et al, 2016). This distribution increases data availability. Also, more data transferred also results in higher costs in used bandwidth. However, comparing the costs of setting up a local data warehouse to using cloud-based services clearly favors cloud computing. And further, security concerns are even more serious when using cloud services, even more so when the externally stored data contains sensitive information (Coleman et al, 2016). An in-house solution which respects the indicators of data security and Legal Regulations would not suffer from the same issue.

For SMEs specifically, a pay-as-you-use pricing model is very attractive. Such pay-per-use solutions are offered by technology leaders such as Amazon Web Services, Microsoft Azure and Google Cloud DataLab (Coleman et al, 2016). The pricing on these offers is flexible and determined by the consumed cloud resources (Coleman et al, 2016). Amazon Web Services depends its pricing on the number of tasks and the data amount scanned in each task. To give an example, 25 tasks a day with each task scanning 1 TB of data would cost USD 3,711 per month (AWS Pricing Calculator, 2021). Whichever provider is chosen, it must be ensured that data security, data ownership protection and a consolidation from multiple sources to a single source of truth is provided (Kalan & Ünalir, 2016).

5.4.2 External solutions

Schuilling (2020) recommends SMEs to engage companies which provide cloud-based analytical solutions to overcome technical barriers when adopting BD. Through that, the challenge of IT infrastructure can be overcome, as it provides a cost-effective solution to engage in analytics (Schuilling, 2020). Additionally, these services often provide human expertise for their customers, mitigating the challenge of BD expertise (Schuilling, 2020).

The reviewed articles mostly discourage the use of external service providers. As most SMEs are proclaimed to only have access to operational level analysis, they face a lack of management analytics (Rajabion, 2018). The often-missing management expertise in SMEs due to their shorter organizational hierarchy further leads to adoption problems when designing, establishing and monitoring a data analytics unit in-house (Coleman et al, 2016). This often forces them to use the costly services of external analysis companies (Rajabion, 2018). Dittert et al (2017; p.3) mentions that external consulting in analytics is *"very expensive"* and mentions the *"high expenses for external experts"*. He relates this towards the slow adaptation of BD in SMEs, as due to their limited financial resources they are not able to take advantage of it. Iqbal et al (2018) further elevates this argument by stating that Data Analytics Consulting Services are usually large firms and as such they cannot align their business practice towards the limited capabilities of SMEs. Further, such consultancy firms sell large teams to their clients which usually solve complex problems for their clients over extended periods of time. And again, he states that such practices are not affordable by SMEs. Resorting to such services also means a loss of control over data (Coleman et al, 2016). This can result in costly lawsuits for an IOSME if e.g. the cloud-service faces a security breach that compromises collected human data of the IOSME. While IOSMEs would of course also benefit from the knowledge of these third parties, it also creates a high dependability.

One key aspect to overcome the challenges is to examine the analytical question an IOSME wants to solve beforehand in order to determine the necessary resources which must be inhibited. Nevertheless, an external solution requires an IOSME to construct the analytical question before commissioning the development as well (Mbassegue, et al, 2016). Setting up an on-site IT infrastructure is costly but also comes with the benefit of more freedom when experimenting with the analytics questions an IOSME wants to answer. Nevertheless, they have to do so at a constant expense (Mbassegue, et al, 2016). Thus, an IOSME has to decide if the

commitment to implement BD internally ‘costs’ more than the benefit they would achieve through external service providers.

To summarize, the tools suitable for BD adoption in IOSMEs were cloud computing and open-source software, represented by Hadoop. These tools are able to mitigate some of the mentioned challenges with respect to the financial restrictions of IOSMEs. Concerning external solutions, the reviewed literature mainly suggests to not engage external service providers due to a higher dependability, costs and them not catering towards the needs of SMEs. However, it has to be noted that the literature which argues against the utilization of external solutions is generally older than the literature arguing for it.

5.4.2.1 Resource Dependency Theory

The Resource Dependence Theory (RDT) was created by Pfeffer & Salancik (1978). It examines how organizations depend on their external environment to obtain critical resources and how this dependence changes organizational behavior. Critical resources are defined as being crucial for the future survivability of the organization. If a company lacks such a critical resource, it will develop relationships with other organizations in order to obtain it (Hazen et al, 2016). Frąckiewicz-Wronka & Szymaniec (2012) state in relation to competitive advantage, that through co-optation organizations can support central tasks through acquiring external resources. This is done by minimizing the dependence towards external actors and maximizing the dependence from the external environment towards the organization. However, this is more relevant towards public companies. Additionally, the RDT has relevance in supply chain management in relation to control over resources that could provide a company with competitive advantage throughout the chain (Delke, 2015). Further, this can be achieved through strategic alliances or merger and acquisitions of companies in the chain.

If the resource is either important, control over it is concentrated or both, it leads to a dependence towards the company controlling such a critical resource (Hazen et al, 2016). This dependence, or power, influences the behavior of involved organizational actors. Such behavior includes actions to counter-act the power an organization has over another. Thus, depending on the particular dependency situation, the RDT is able to explain organizational decisions and actions. Organizational success in RDT is explained as organizations efforts to maximize their power (Nemati et al, 2010). Thus, organizations actively try to reduce their dependency towards other organizations.

In relation to BD, the RDT was mainly used towards supply chain management in academic research (Prasad et al, 2018). In the case of humanitarian supply chain networks, it was noted that their dependence towards governments and donor agencies can lead to a widespread adoption of BD (Prasad et al, 2018). This is due to the strong interdependence between external stakeholders and NGOs in e.g. cases after a natural disaster, where deliveries of supplies are relying on them. There, high expectations from governmental-side can shape the activities of the NGOs towards BDA utilization in order to ensure an efficient process (Prasad et al, 2018).

5.5 Conceptual framework

Through the findings in the literature, a conceptual framework is created to provide an overview on how the available options can be scaled according to their investment. This framework provides important guidelines for strategic management in an internationally operating SME.

According to the examined literature, the first step when deciding for an incorporation of BD into business practice is to define the analytical problem an IOSME wants to solve. This scope of the analytics process is influenced by various factors, e.g. which data is utilized, which tools will be used in the BDA process, the complementing IT infrastructure, the available BD expertise and if the company is able to successfully consume the benefits. The choice of consulting external service providers was highly disregarded in the examined literature due to its cost and them not being aligned to the needs of IOSMEs. Additionally, according to the RBV, a sustainable competitive advantage cannot be reached through employing external resources. Thus, the internal solutions for BD implementation were considered as Cloud Computing and open-source software, with specific focus towards Hadoop. Depending if human or non-human data is utilized in the process, the existing challenges for BD adoption differ. Namely the Legal Regulations and Cross-cultural barriers are to be considered if the IOSME wants to conduct BDA on human-generated data. Through the suggested solution towards each challenge these can be overcome, which enables companies to consume the benefits, generated through the insights which were uncovered which in turn support the firm's decision-making process. As mentioned before, the investment into a BD infrastructure itself leads not to a competitive advantage. To reach that point, the BDA capabilities need to be aligned towards the properties of the VRIO-framework, namely valuable, rare, costly to imitate and organizational value-capturing. By inheriting these features into the BDA process, they are

able to achieve a sustained competitive advantage, which in turn enhances their survivability and competitiveness. The derived conceptual framework is depicted in Figure 13.

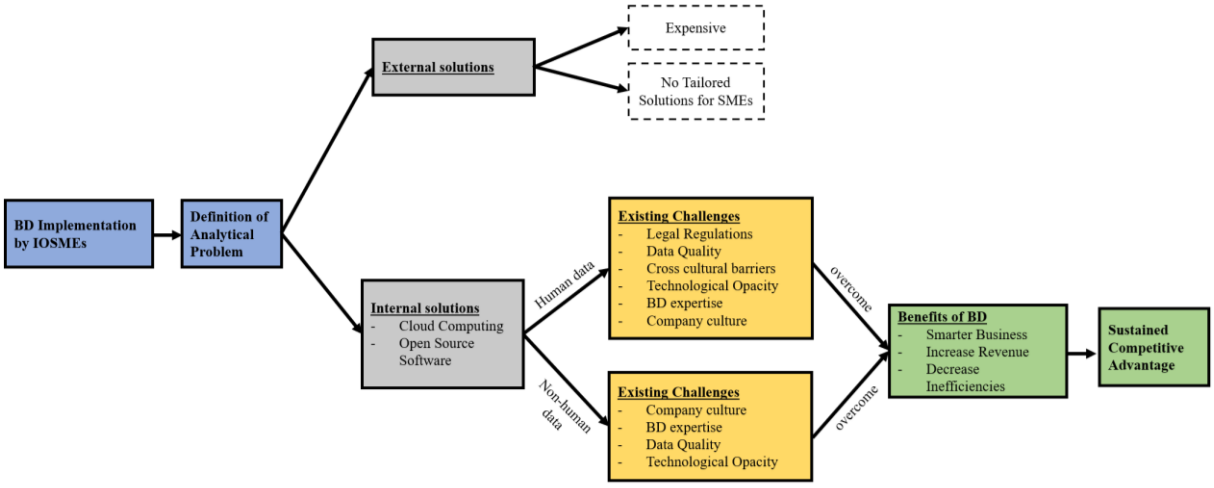


Figure 13: Conceptual framework

own compilation

6. Empirical Analysis

It was proposed in the literature that there are currently no solutions from external providers available with regard to the requirements of SMEs or IOSMEs. The empirical analysis which is based on secondary data will shed light on whether there are possible solutions available and if the literature is thus outdated. By searching for suitable data sources, Transmetrics and AtScale have found to be suitable examples to illustrate that newly emerged businesses indeed have developed solutions that are sufficient for the needs of IOSMEs. Additionally, these companies are IOSMEs. They were specifically selected in regards to the mentioned statement in the literature review, that external service providers are not aligned towards the needs of IOSMEs. Being an IOSME themselves, they inherit the necessary knowledge to align their offering towards IOSME-customers with their specific limitations. The companies will be analyzed independently based on the fact whether the analytical problem is concerned with human or non-human data. The empirical analysis will then be used to extend the current knowledge within the literature.

The initial coding relies on the findings from the literature review. The used categories in the initial coding phase are visualized in Figure 14.

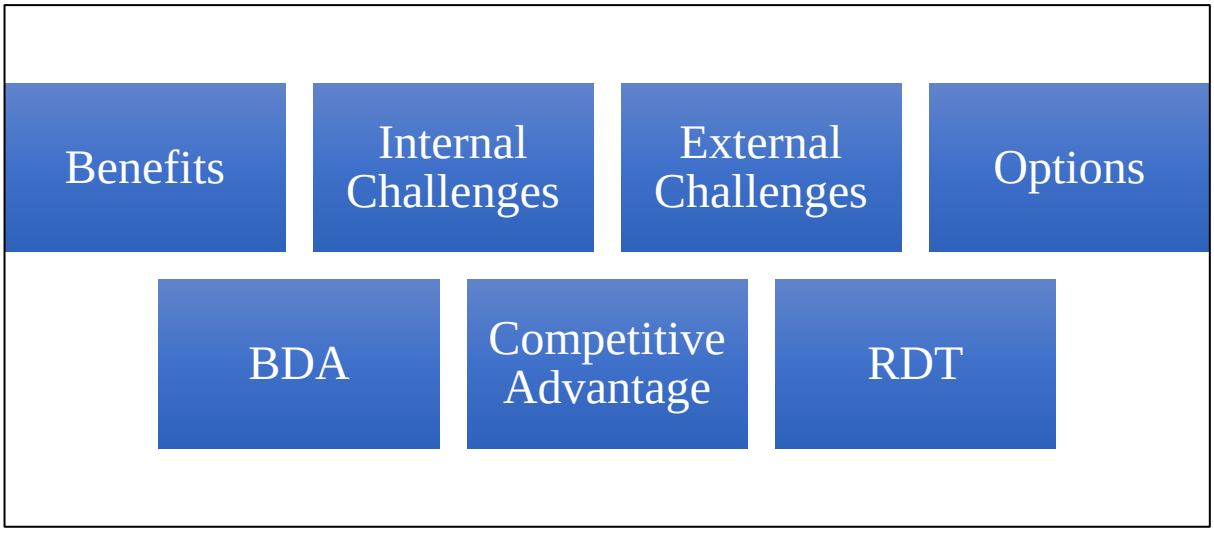


Figure 14: Used initial coding categories

own compilation

6.1 Data Analysis of Case Studies

Transmetrics

In this section, the case of Transmetrics is examined. Firstly, the business processes of Transmetrics will be explained in detail. Thereafter, the processes are related towards the contents of the literature review to provide an overview.

Transmetrics offers their Software-as-a-Service solution, where all hardware and software is provided, through a monthly subscription basis (Analyze - Transmetrics, 2021). When using their external service, their first step is to connect to the enterprise resource planning and asset management tools used by their customers in order to gain access to their data (Transmetrics.ai, 2021). This data is then extracted by an automated process on a daily basis and stored on a cloud computing database (Transmetrics.ai, 2021). The transport management software typically used by logistic companies, is often run on terminal based mainframes with limited computing power and data storage (Transmetrics - A cloud solution, 2021). Transmetrics liberates this local data by moving it to the cloud. After having gained access to the data, they deploy a machine learning algorithm to reduce contained noise (Transmetrics.ai, 2021). After the transformation of data, they use BDA to provide their customers with an in-depth analysis

of historical operational performance, trends and most importantly, identified inefficiencies (Transmetrics.ai, 2021). This is achieved through combining customer data and external data sources such as port data or weather logs. Through that, they are able to construct a foundational database through which predictive analytics can be conducted (Analyze - Transmetrics, 2021). In the case of a customer they managed to attain a 96% accuracy in their logistics predictions (Koev, 2021). Transmetrics states that a major selling point in integrating their BDA software is to shape the pool of data in a company towards a single version of truth (Predictive Empty Container Management, 2021). They further state that there is a need to move planning from individual people towards an IT system which supports complex operational decisions. Before integration, forecasts are forged by an internal process depending on manual inputs and personal business experience (Predictive Planning - Case Study, 2021). Generally, their service requires customers to submit at least 6 months of historical data in order to sufficiently provide BDA (Analyze - Transmetrics, 2021). Their solution for new customers that were not able to provide such data was to enable users of their customers to manually suggest their predictions due to no previous data being available (Predictive Planning - Case Study, 2021). While this produced satisfactory results, it is not optimal in the long-term due to it not relying on statistical measurements. In a case project to support predictive empty container management for their customer, they set up a shared team of project managers, IT personnel, business users, business analysts and data scientists (Predictive Empty Container Management, 2021). This enables them to focus their human resources mainly on their offered service, while the expertise and company insights are provided by the team of the customer. The services of Transmetrics are priced at EUR 3,500 to EUR 5,000 per month ("Transmetrics Project", 2021). Additionally, they offer try-before-you-buy packages of their software ("CORDIS EU", 2021).

In regard to Benefits, the end-result of their BDA process relates to smarter business by reducing the number of workers needed to transport the products of their customers. Further, by lowering the empty space in the used form transport they provide their customers with the benefit of decreasing inefficiencies. However, the use of Transmetric’s services does not lead to an increase in revenue, as their service is not directed at enlarging customer bases.

As the provided cloud database can be accessed by their customers through any web browser, Transmetrics mitigates the internal challenge of setting up a own IT infrastructure. However, they transfer only logistics-related data to the cloud. Thus, a customer still has to adapt towards an IT infrastructure that is able to comprise BD. But as cloud computing is a platform that can

be easily scaled towards the needs of IOSMEs, the challenge can be seen as being overcome. Their advances towards a supportive decision-making tool that has less need for business experience can lead to opposition in their customers company, elevating the challenge of Company Culture for them. This aspect is crucial as Transmetrics relies on shared teams where efficiency might suffer from non-supportive employees. But as the usage of their services produces immediate beneficial business suggestions that can be implemented, it reduces the uncertainty towards BD as a new technology. However, this challenge might be mitigated through the fact that the implementation process of BD relies mostly on Transmetrics. Thus, it can be argued that the transformation of the existing organizational culture towards a BD embracing company is not as crucial as would be the case when BD is implemented internally. Thus, the challenge of Company Culture is met. When viewing BD expertise as an internal challenge, it becomes clear that Transmetrics does not support customers in overcoming it, as they inherit the whole BD expertise. Although there might be slight insights in the shared teams, a IOSME does not obtain BDA capabilities through the use of their services.

Transmetrics engages in the External Challenge of Data Quality in multiple ways. The Availability is ensured as they grant Accessibility and Timeliness towards the extracted data of their customers terminal mainframes. The Authorization to access this data is ensured through the integration of authentication methods that can connect to the customers IT system ("CORDIS EU", 2021). As the logistics industry is heavily reliant on human input, a missing common standard for logistics processes often results in incomplete datasets. The data-transforming machine learning algorithm thus further enhances Reliability. Lastly, the usage of shared teams increases Usability as they bridge their own BD expertise and their customers business knowledge. The missing aspects of Relevance and Presentation Quality in regard to the challenge of Data Quality are comprised in the usage of their service but might not apply in the context of using external service providers. When deciding if the challenge of Data Quality it has to be noted that Transmetrics provides the cleaned data towards their customers as well. Thus, the challenge of Data Quality is met entirely. With the utilization of non-human data, the external challenge of Legal Regulations plays a lesser role. This is also true for the challenge of Cross-Cultural barriers. However, Transmetrics does not provide any support towards these challenges. Going further, Transmetrics takes advantage of shared teams. This provides employees from the customer side with vital insights in how certain BDA processes function. However, it can be assumed that Transmetrics does not provide guidance for their customers in

regard to which BDA tools to utilize, as this would damage their own business. Thus, the challenge of Technological Opacity is not met.

Transmetrics uses the Option of Cloud Computing. The sensitive topic of data security when engaging in cloud computing is ensured by various layers of security. The connection through which the customer data is extracted and stored into the cloud database is established through a virtual private network (Analyze - Transmetrics, 2021). Through that, they establish an encrypted connection to transfer the data between devices (VPN usage, 2021). This makes it nearly impossible for the data to be breached, as it would take either the utilized encryption key or to manually re-translate the data. Both of which would take an immense financial effort that outweighs the worth of the stolen data. And further, Transmetrics is ISO 27001 certified, which is an international standard ensuring information security (Analyze - Transmetrics, 2021). In detail, this standard ensures that the company assesses the potential risk when dealing with data from third parties, takes the necessary steps to defend against them and continuously improves their information security management system (ISO 27001 Requirements, 2021). Through that, they ensure that the transferred data is well secured from security breaches. The conversion towards a single version of truth was mentioned as a requirement for cloud computing in the literature review, as having multiple databases, also called ‘silos’, can lead to differing realities when conducting BDA.

Regarding the category of BDA, it can be argued that the customer data by itself might not fall into the category of BD, in regard to Variety and Volume. But by combining different data sources in addition to the customer data Transmetrics enrichens the analyzed data towards more accurate results. Although, this additional data sources are not shared with their customers, which makes them unable to conduct BDA through the external service.

Transmetrics achieves a competitive advantage against their competitors by combining BDA and a difficult to imitate process of collecting, preparing and analyzing data. Although their customers are able to achieve the stated benefits, they will not be able to obtain a sustainable competitive advantage through their services. According to the VRIO framework, only temporary competitive advantage can be reached as the service can be used by their competitors as well.

When viewing their services through the RDT, it becomes clear that a utilization of their services results in a high dependence as Transmetrics conducts the whole BDA process by themselves. This dependence might be increased even more if the competition starts utilizing their services as well. However, this is dependence is mitigated by their service relating to a monthly subscription service. Thus, customers are not bound to a long-lasting contract. It can be assumed that IOSMEs who utilizes their services realize the need to obtain BDA capabilities according to the VRIO framework in order to reach a sustainable competitive advantage. This can be seen as a counter-activity in order to reduce their dependence towards Transmetrics. Nevertheless, the dependency exists when engaging in their services.

AtScale

AtScale provides a virtual datacenter based on cloud computing that enables their customers to optimize their own analytics process. When implementing their services, an online data warehouse based on Amazon Web Service is set up which operates with AtScale's software ("AtScale Case", 2021). The company then transfers their data towards this data warehouse and connects their analytical software with it. AtScale achieved to differentiate itself from its competitors by creating a 'no-data-movement' architecture where the data is virtual and not materialized as it would be when it is stored in a server ("AtScale Hadoop", 2021). Thus, the data is provided immediately per request. Users are then able to access all data at once and not just pre-aggregated results like it would be with traditional storing methods. Secondly, data modelling can be conducted without extracting the data first ("AtScale Hadoop", 2021). Thirdly, the data is simultaneously accessible by users ("AtScale Hadoop", 2021). Additionally, they apply autonomous data engineering, which includes transformation, maintaining and delivery of data ("AtScale Solutions", 2021). AtScale's mission is stated as bridging the gap between existing Analytics ecosystems and Hadoop (AtScale on Hadoop, 2021). They achieve that by utilizing OLAP (Online Analytical Processing) servers to store data. Through that, traditionally used tools such as Excel, which would normally not be able to handle such amounts of data properly due to their processing limits, are able to interact with the Hadoop platform. AtScale ensures Data Security in their cloud computing-based architecture through authorization procedures for users and object level security which secures access towards sensitive data only for specific users ("AtScale Security", 2021). AtScale further adds a semantic layer between Hadoop and the tools used to gain insights. BD normally consists of large, unstructured data which needs to be filtered, cleaned and maintained before conducting BDA. Through the semantic layer, data is segmented automatically. This allows customers to

choose fluently between data categories, e.g. products, revenue or other consolidated views of data ("AtScale on Hadoop", 2021). AtScale mentions a use-case of their software for the shopping rewards company Rakuten Rewards, headquartered in the US. Their customer initially had an on-site Hadoop cluster in order to conduct BDA on customer data ("AtScale Case", 2021). After being initially successful with their undertaking, they gradually experienced business disruptions due to processing issues ("AtScale Case", 2021). While Hadoop is intended to simultaneously process user queries, their IT infrastructure was overloaded due to spikes in demand at the beginning and the end of each month. Namely, the marketing department ran large-scale email campaigns while the finance department conducted month-end close calculations ("AtScale Case", 2021). With 1.5 million monthly queries the scalability of the system was not feasible for a longer period of time ("AtScale Case", 2021). Thus, they changed their on-premises Hadoop cluster towards a cloud data warehouse with AtScale. This resulted in faster conducted reports without any service interruptions while they benefited from the compatibility of AtScale's OLAP architecture ("AtScale Case", 2021).

They achieve the benefit of Smarter Business as their platform enables users to access all historical data at once, which leads to more reliable insights. This in turn strengthens the analytics process of their customers as a whole, further supporting the decision-making process. Further, their services increase the efficiency of conducted BDA processes, as data does not need to be extracted or transformed in a separate step. Due to that, existing inefficiencies in workforce and time spent on analytical processes are decreased. However, as AtScale does not provide insights on the customer data itself, they do not provide the benefit of increased revenue.

AtScale offers a virtual IT infrastructure for their customers which solves the internal challenge of IT Infrastructure for them. It further can be argued that the internal challenge of Company Culture is mitigated as their platform supports traditional analytic tools. Due to that, the number of opposing mindsets toward BD implementation decrease. Nevertheless, the customer organization still needs to ensure a BD embracing company culture in order to reap the benefits of BD implementation. Thus, the internal challenge of Company Culture remains. In a normal BDA setting, IT employees are tasked to extract the data needed in the BDA process from on-site servers. The service of AtScale overcomes this task by enabling access through their no-data-movement architecture. Subsequently, it positively impacts the necessary BD employment related to this function. This is further enhanced by their autonomous data engineering process

and their semantic layer, as it gets the data ready for analysis automatically. While AtScale does not offer BD expertise towards their customers, as was the case with Transmetrics, their compatibility with non-BD tools reduces the challenge of BD expertise. It relates to a subsequent lower need for experts that inherit knowledge of specialized BD software in order to handle the amount of data analyzed. Depending on the scope of the analytics process used in the IOSME, this challenge can be overcome entirely. Considering this, one can argue that the challenge of BD expertise is overcome.

Regarding the external challenge of Data Quality, it is ensured through authorization procedures and simultaneous access of users towards the data, enhancing Availability. The Reliability of Data Quality is further strengthened through their autonomous data engineering process. The semantic layer further enhances Relevance, as the data is already segmented towards data categories, ensuring that the data fits the theme of analysis. Considering that existing analytical processes already inherit the necessary Usability and Presentation Quality, the challenge of Data Quality is met. When examining the external challenge of Legal Regulations, AtScale provides several layers of security to ensure Data Security for the customer data. Thus, human-related data can be used and analyzed through their services. However, AtScale does not provide legal guidance towards obtaining and using data from external resources. Subsequently, the challenge of Legal Regulations is not met. This also applies towards the challenge of Cross-Cultural Barriers.

As the overall goal of using external service providers is to reach a sustainable competitive advantage, it must be examined if the analytical process fulfills the requirements of the VRIO framework. But by enabling non-BD tools to interact with a data warehouse that is able to handle BD, the challenge of Technological Opacity can be regarded as resolved.

The utilized Options of AtScale are Hadoop and Cloud Computing. Their product overcomes the drawbacks of Hadoop while maintaining the advantages of this open-source software. AtScale's virtual datacenter is based on a cloud computing platform. Their platform allows for isolated data centers across companies, the aforementioned silos, to be transformed into a single virtual data warehouse. Again, the aim of this process is to shape the data architecture towards a single source of truth. As such, it eliminates the need for data movement, hardware or software changes or separate clusters for storing data. This mitigates the costs when scaling the existing BD infrastructure, as e.g. no additional servers for storing data need to be bought. Instead, the

computational capacity can be fluently extended by choosing a different service plan. While this would also be possible with a traditional cloud computing service, the key difference is the compatibility with Hadoop.

Concerning BDA, AtScale's semantic layer and data engineering engages in data transforming processes. However, they do not conduct any analytical processes that could support decision-making. Further, the utilization of their services enables customers to utilize non-BDA tools to engage in BDA. This has to be critically assessed as tools such as Excel or Tableau cannot compute complex BDA. However, as it enables these tools to use BD as basis for their analytical process, it enhances the efficiency of their customers BDA. Further, the analytics process can inspect more data without enhancing its duration, while the resulting insights are uncovered faster and produce more trustworthy results, as more data strengthens the achieved insights. Through the semantic layer, the BDA process is further enhanced in efficiency as it automatically segments the data. And further, as non-BDA tools normally would not be able to handle such data segmentation tasks, the semantic layer enables them to only focus on needed data segments. This reduces the task time of the BDA process again. By enabling their customers to interact with the data simultaneously, it leads to a higher efficiency in the BDA process. However, while their service enhances the BDA process in numerous ways, AtScale does not provide support towards building BDA capabilities within the organization of their customers.

Using non-BD tools limits the BDA process towards less complex methods. This does not rule out an achievable sustainable competitive advantage, which is depending on existing BDA capabilities that apply to the VRIO framework. While the analytical process is strengthened due to more data being available for analysis, it can be generally assumed that the traditional tools do not use processes that are imitable for competitors. Thus, using the external service of AtScale does not lead to competitive advantage.

With relocating all data towards a virtual architecture, customers become dependent on AtScale's service. Although, the stored data in the cloud can be transferred towards other cloud providers or on-site data warehouses, the customer is reliant on AtScale's product as to not interrupt business processes.

6.2 Comparison of Cases

Both cases offer their service in order to achieve a shared goal: advance BD implementation in smaller- and medium-sized businesses. While this is realized through their offered products, they share similarities and differences in achieving it. The most prominent difference between the cases is that Transmetrics actively engages in BDA in order to provide solutions towards their customers, whereas AtScale does not. However, this only translates to differences in overcome challenges but has no effect on the achieved benefits. Through examining the case studies it became clear that while some challenges can be mitigated through external services, some challenges remain to be overcome by IOSMEs in their entirety. The overall goal of implementing BD is to reach a sustainable competitive advantage through building own BDA capabilities that inherit the aspects of the VRIO framework. The selected cases showed that while the BDA process improves through their services, they do not support building of such capabilities for their customers. This can be explained through the RDT, as such advantages would result in a loss of power for such services. Furthermore, both services lead to an increase in dependency for their customers. The similarities and differences are summarized in the table visualized in Table 3.

Table 3: Summary of case comparison

		Transmetrics	AtScale
Benefits	Smarter Business	✓	✓
	Decrease Inefficiencies	✓	✓
	Increase Revenue		
Internal Challenges	IT Infrastructure	✓	✓
	BD Expertise		✓
	Company Culture	✓	
External Challenges	Data Quality	✓	✓
	Legal Regulations		
	Technological Opacity		✓
	Cross-Cultural barriers		
Options	Hadoop		✓
	Cloud Computing	✓	✓
BDA capabilities			
Competitive Advantage			
Resource dependency		✓	✓

6.3 Axial Coding

The previous category of Internal Challenges has differing implications towards the Phenomenon of slow BD adaptation in IOSMEs and SMEs. Therefore, the internal challenges were split towards attributes that influence organizational strategy towards BD implementation and underlying reasons that led to this Phenomenon. Accordingly, the category of Causal Conditions now contains the internal challenge of IT Infrastructure, as it was examined to be a primary reason for slow BD adoption. The other internal challenge of Company Culture can be viewed as more influential towards organizational strategy, thus falling into the category of Intervening Conditions. BD expertise, being regarded as internal and external challenge, is viewed as an internal challenge in this framework as it is more difficult to obtain external BD expertise when compared with building internal BDA capabilities. Subsequently, the challenge of BD expertise is also categorized as a Causal Condition for slow BD adoption. The new

framework with the adapted categories is visualized in Figure 15. This adapted framework is then used for the cases of Transmetrics and AtScale based on which challenges were overcome and which categories they provide through their services. For the case of Transmetrics, the adapted framework can be seen in Appendix 1, while AtScale's framework is visualized in Appendix 2. Ultimately, the last step of selective coding enhances the conceptual framework with the insights gathered in the empirical analysis. This results in the adapted conceptual framework which will be provided in the following section.

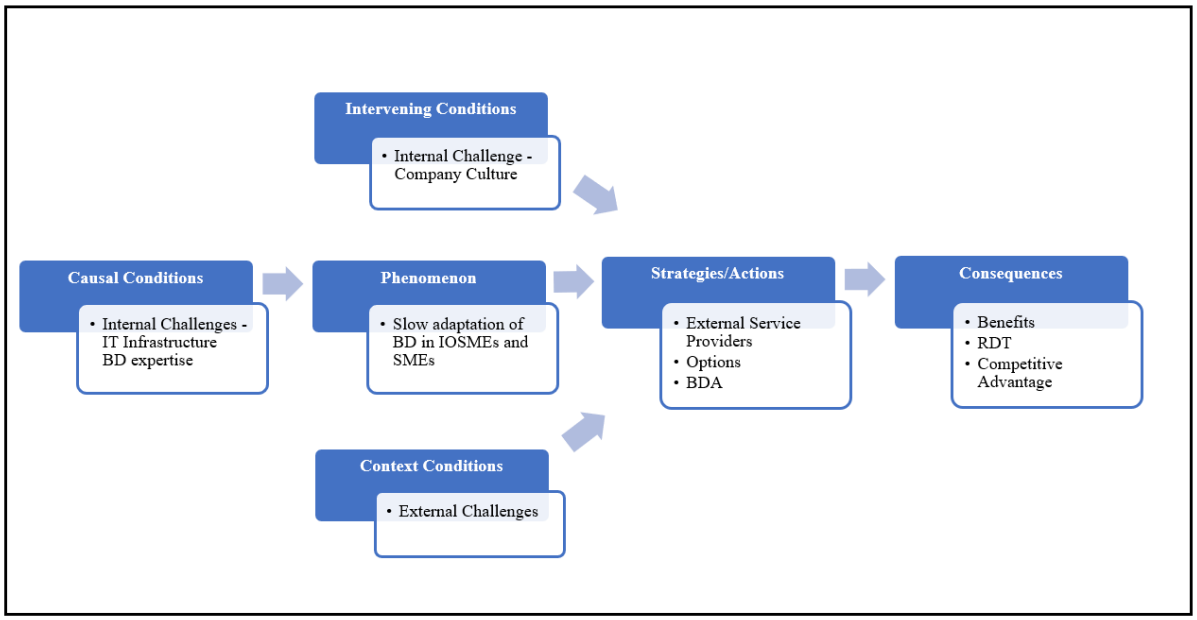


Figure 15: Adapted Axial framework

own compilation, based on Corbin & Strauss (1990)

7. Discussion

Academic articles in the literature review stated that the use of external services for BD implementation in SMEs is not feasible due to service providers not being aligned towards the needs of SMEs and their expensiveness. However, the empirical analysis showed that service providers, which aligned their products towards SMEs and consider their limited financial capabilities, do in fact exist. This can be explained with the lowering prices for computational power and data storage due to technological advancements and the demand for such easy-to-integrate options due to the current slow adaptation.

The literature showed that a sole investment into a BD infrastructure does not lead to a competitive advantage. Such an advantage can only be achieved by utilizing BDA capabilities

that inherit capabilities according to the VRIO framework. This was captured in the challenge of BD expertise which is hard to overcome for IOSMEs due to the lack of available labor, the need of 'all-rounders' and the high salaries for such experts. This leaves IOSMEs with the decision to either upskill their current team or to use external service providers. Such upskilling is further affected by the challenges of Company Culture and Technological Opacity, which makes the option external providers more attractive. The comparative case study showed that external providers help overcome the challenges of IT infrastructure and Data Quality in order to efficiently provide their own services. Depending on the service used, additional challenges can be overcome as well. How extensive their support towards these challenges can be relies on the service they perform, whether they engage in BDA themselves or not, and the scope of the analytics process in the IOSME. Overcoming initial challenges makes the utilization of external service providers an ideal starting point towards BD implementation. The consideration of external solutions is further supported by pricing models that regard the limited financial capabilities of IOSMEs with e.g. test-before-you-pay offerings. However, such services can be used by competitors as well, mitigating the achievable competitive advantage. Additionally, the examined external providers do not support the enhancement of BDA capabilities in their customers organizations. This is crucial, as only through building such BDA capabilities that inherit valuable, rare, inimitable and organizationally-capturable properties, a sustainable competitive advantage can be reached.

This is in line with the RBV, as only internal resources can lead to a competitive advantage. Considering the applicability of the VRIO framework in the case of BD and its subsequent BDA process as a resource, it can be seen as valuable as the benefits ultimately lead to decreased costs by reducing inefficiencies or to differentiated products that satisfies unmet customer preferences. By arguing that BD is only slowly adopted by companies due its complexity and costs, one can argue for BD being a rare tool. Nevertheless, the rarity of BD utilization highly depends on the industry and the countries an IOSME operates in. If an IOSME would be one of the first to incorporate BD in its industry, its competitive advantage would be unmistakably higher than in industries where BD is common practice. It can be assumed that a thorough analysis of the tools used by current competitors might bring insights in how to outperform them in their BDA capabilities and thus, achieve a competitive advantage. If an IOSME specializes towards an in-depth analysis according to their industry by using a complex mix of BDA processes and capabilities and data which is hard to access, they ensure that their BD resource is inimitable. The point of organizing the company structure to capture the value of

BD relates to the Company Culture. Subsequently, an IOSME can reach sustainable competitive advantage through the use of BD. However, this perspective is challenged by the RDT that claims a competitive advantage can also be reached by forming strong alliances that share heterogeneous resources. Although BDA capabilities can be seen as a resource in the RBV, it can be argued that the competitive advantage through merger and acquisitions or strategic alliances, in the context of the RBV, might not be applicable towards IOSMEs due to their limited financial abilities. Further, as the BDA process that ensures a sustainable competitive advantage might be easily replicated if performed on obtainable datasets, a company that withholds such BDA capabilities would disregard sharing it with other companies in order to not lose the advantage. To answer the overall RQ and considering that a successful BD implementation should ultimately lead to a sustainable competitive advantage, it can be concluded that it cannot be achieved through external service providers. While external service providers improve the overall BDA process for their customers, they do not build up BDA capabilities in organizations of their customers.

With regards to the RDT, it can be seen that the use of external services creates interdependencies between IOSMEs and external service providers, with IOSMEs becoming dependent on the offered services in order to reap the benefits and external service providers relying on the financial resources in order to ensure their own organizational survivability due to them being IOSMEs as well. Such interdependencies lead to ‘power struggles’ with both parties engaging in counter-acting strategies to reduce such dependence. In the case of IOSMEs such actions could include the training towards own BDA capabilities with the goal of reaching a sustainable competitive advantage by themselves. Considering that a IOSME is able to benefit from external services that support BD implementation while overcoming challenges of BD adoption while counteracting a growing dependence through building own BDA capabilities with the aim of reaching a sustainable competitive advantage, the utilization of external service providers for that cause can be viewed as beneficial. Through the reinvestment of the achieved benefits through external solutions, an IOSME can gradually build up their own BDA capabilities to, in the end, reach a sustainable competitive advantage on its own. This is visualized in Figure 12, the adapted conceptual framework. This makes external solutions a prosperous ‘stepping stone’ towards BD implementation, which leaves IOSMEs to determine if the achievable benefits outweigh the necessary effort that comes with BD implementation.

As the overall goal of BD implementation is to reach a sustainable competitive advantage, it is hereby identified as the core theory. The final output of the grounded theory approach resulted in the theoretical framework, which is visualized in Figure 16. The BD implementation process starts in the pre-adaptation phase by an IOSME identifying the related challenges. According to the literature review, these challenges vary depending on multiple internal and external constraints e.g. which type data is used, the scale and scope of the BDA process and Cross-Cultural Barriers. Then, the IOSME can decide whether to overcome these challenges themselves or if they want to utilize external services. In the case of overcoming the challenges on its own, beneficial options that support the limited capabilities of IOSMEs should be considered. They further support a IOSME in the next step of overcoming the identified challenges. Having successfully reached the post-adaptation phase, the IOSME is able to reap the benefits of their efforts. Through building internal BDA capabilities that fulfill the requirements of the VRIO framework an IOSME is able to reach the goal of sustainable competitive advantage. In the case of utilizing external services, the benefits are achieved faster but still require the IOSME to overcome challenges that are not solved through such services. In order to counter the resource dependency towards their external service provider, a IOSME should then engage in building own BDA capabilities, again with the overall goal of reaching sustainable competitive advantage.

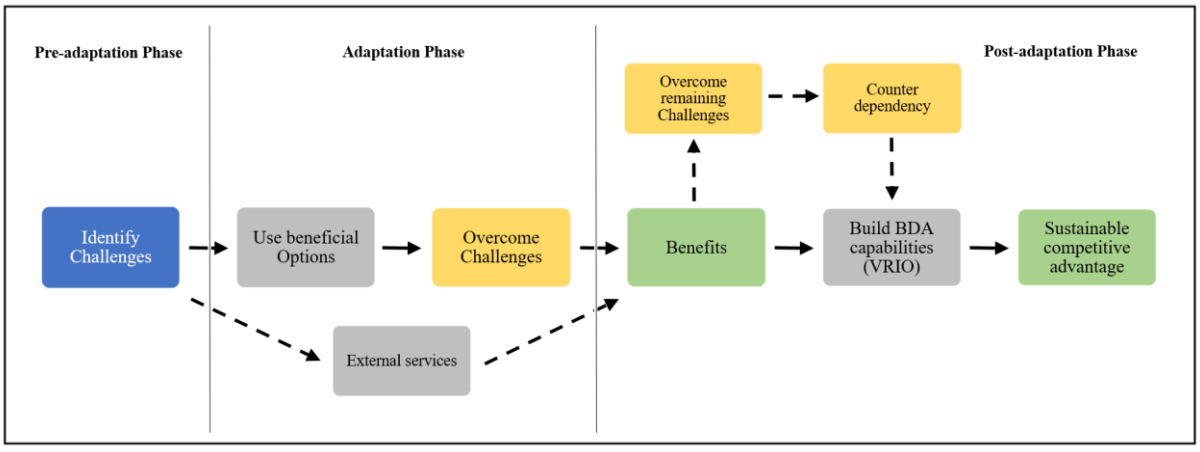


Figure 16: Adapted Conceptual framework

own compilation

8. Limitations and Future Research

The Limitations are concerned with occurred issues during the research process and their influence on this thesis. Starting off, the case study part of this research was only conducted on

two case companies. Their offering for customers is highly specialized. Thus, conducting a case study with different case companies might bring differing results and insights towards the research topic. Further, the field of research concerning Big Data and Internationalization is subject to fast technological advances. This makes it necessary to view it in the lights of the current technological achievements which subsequently leads increases the risk of the attained knowledge in this thesis in being outdated over a shorter period of time. The overall aim of this thesis related to a very generalized approach towards BD implementation in IOSMEs. Narrowing the scope of this project towards specific industries, countries or businesses can lead to more fitting insights towards companies in the chosen sphere.

With research of BD integration in IOSMEs being highly underdeveloped, future research that revolves around the usage of external resources in order to achieve a competitive advantage would further advance academic knowledge in this scientific area. Literature already investigated the role of strategic alliances in achieving sustainable competitive advantage. It can be implied that if only one resource is attained through an external resource while the bundle of resources still obtains the properties according to the VRIO-framework, a sustainable competitive advantage can be achieved, given that the bundle of resources is unique. Additionally, future research can examine how BDA capabilities can lead to sustainable competitive advantage in detail with respect to contextual matters such as industry, business model, etc. Future research would also benefit from a subjectivistic approach to investigate the challenge of Company Culture. Through determining the motivations which lead to a retrogressive towards BD implementation, new insights can be uncovered which in turn can support IOSMEs with their BD implementation efforts.

9. Conclusion

The conducted literature review gave a broad understanding over challenges and benefits of BD implementation for IOSMEs. Further, beneficial tools which are beneficial in regard to the limited resources of IOSMEs were examined in accordance with their implications on such challenges and benefits. Examined articles in the literature review generally advised against the use of external service providers due to them not being too expensive and not aligned with the needs of IOSMEs. However, the comparative case study showed that such external services do in fact exist. It further showed that such services could provide a ‘stepping stone’ towards BD adoption in IOSMEs. This is due to external service providers support overcoming several initial challenges that arise from BD implementation. However, this comes with the price of a

high dependence towards these service providers. The examined cases showed that while these providers enhance efficiency in the BDA process of their customers, they do not support building BDA capabilities that inherit the capabilities of the VRIO framework. Such BDA capabilities are crucial towards reaching a sustainable competitive advantage. Thus, a IOSME cannot reach a sustainable competitive advantage through the use of external services alone. By using the suggested recommendation of ‘think big but start small’ these IOSMEs can then gradually intensify their advances towards their own BDA capabilities to counteract the dependence towards these external service providers and ultimately achieve a sustainable competitive advantage.

10. References

Agarwal, R., & Dhar, V. (2014). Editorial—BD, Data Science, and Analytics: The Opportunity and Challenge for IS Research. *Information Systems Research, 25*(3), 443-448. doi: 10.1287/isre.2014.0546

Alhojailan, M. I. (2012). Thematic analysis: A critical review of its process and evaluation. *West East Journal of Social Sciences, 1*(1), 39-47.

Analyze - Transmetrics. (2021). Retrieved 17 April 2021, from https://www.transmetrics.ai/platform/analyze/

Arbnor, I., & Bjerke, B. (2009). *Methodology for creating business knowledge*. Sage.

AtScale Case 2. (2021). Retrieved 31 May 2021, from https://www.atscale.com/resource/global-agriculture-chemical-and-energy-leader-transforms-their-operational/thank-you/

AtScale Case. (2021). Retrieved 19 May 2021, from https://www.atscale.com/resource/case-study-atscale-and-rakuten/

AtScale Company. (2021). Retrieved 24 April 2021, from https://verify.wiki/wiki/AtScale

AtScale Profile. (2021). Retrieved 24 April 2021, from https://craft.co/atscale

AtScale Security. (2021). Retrieved 19 May 2021, from https://www.atscale.com/wp-content/uploads/2019/07/AtScale_HowItsDone_Security.pdf

AtScale Solutions. (2021). Retrieved 24 April 2021, from https://www.atscale.com/#

AWS Pricing Calculator. (2021). Retrieved 14 April 2021, from https://calculator.aws/#/createCalculator/Athena

Barney, J. (1991). Firm resources and sustained competitive advantage. Journal of management, 17(1), 99-120.

BD. (2019). In *Oxford Online Dictionary*. Retrieved 8 March 2021, from https://www.oxfordlearnersdictionaries.com/definition/english/big-data?q=Big+Data

Bell, E. & Bryman, A. (2011). Business Research Methods, 3rd ed., Oxford University Press

Bell, E., Bryman, A., & Harley, B. (2018). *Business research methods*. Oxford university press.

Bertello, A., Ferraris, A., Bresciani, S., & De Bernardi, P. (2020). BDA (BDA) and degree of internationalization: the interplay between governance of BDA infrastructure and BDA capabilities. *Journal of Management and Governance*, 1-21.

Big Data Failure. (2021). Retrieved 11 May 2021, from https://www.leadboxer.com/blog/when-big-data-goes-wrong/

Bryman A. & Bell, E. (2015) "Business Research Methods" 4th edition, Oxford University Press, p.27

Burrell, G., & Morgan, G. Sociological Paradigms and Organisational Analysis, 1979. *London: Heinmann*.

Buy Database. (2021). Retrieved 11 May 2021, from https://mailingdatapro.com/

Cai, L., & Zhu, Y. (2015). The challenges of Data Quality and Data Quality assessment in the big data era. *Data science journal*, *14*.

Cambridge Analytica. (2021). Retrieved 11 May 2021, from https://www.wired.com/amp-stories/cambridge-analytica-explainer/

Cambridge Analytica Aftermath. (2021). Retrieved 11 May 2021, from https://www.bbc.com/news/technology-44807093

Chun Tie, Y., Birks, M., & Francis, K. (2019). Grounded theory research: A design framework for novice researchers. SAGE open medicine, 7, 2050312118822927.

Coleman, S., Göb, R., Manco, G., Pievatolo, A., Tort-Martorell, X., & Reis, M. S. (2016). How can SMEs benefit from BD? Challenges and a path forward. *Quality and Reliability Engineering International*, *32*(6), 2151-2164.

Company Transmetrics. (2021). Retrieved 17 April 2021, from https://www.topionetworks.com/companies/transmetrics-5a1bddd8105eb57b63405404

CORDIS EU. (2021). Retrieved 31 May 2021, from https://cordis.europa.eu/project/id/945610/reporting

Corbin, J., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. Qualitative Sociology, 13, 3-21.

Dam, N. A. K., Le Dinh, T., & Menvielle, W. (2019). A systematic literature review of BD adoption in internationalization. *Journal of Marketing Analytics*, *7*(3), 182-195.

Delke, V. F. (2015). The resource dependence theory: Assessment and evaluation as a contributing theory for supply management (Bachelor's thesis, University of Twente).

Denyer, D., & Tranfield, D. (2009). Producing a systematic review.

Diebold, F.X. (2000), "BD Dynamic Factor Models for Macroeconomic Measurement and Forecasting," Discussion Read to the Eighth World Congress of the Econometric Society, Seattle, August.

Diebold, F. X. (2012). On the Origin (s) and Development of the Term 'BD'.

Dikshantmalidev. (2021). Hadoop - Pros and Cons - GeeksforGeeks. Retrieved 18 May 2021, from https://www.geeksforgeeks.org/hadoop-pros-and-cons/

Dittert, M., Härting, R. C., Reichstein, C., & Bayer, C. (2017, June). A data analytics framework for business in small and medium-sized organizations. In *International conference on intelligent decision technologies* (pp. 169-181). Springer, Cham.

Duschek, S. (2004). Inter-firm resources and sustained competitive advantage. In *Management Revue*, pp. 53-73.

Espinosa, J. A., & Armour, F. (2016, January). The BDA gold rush: a research framework for coordination and governance. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 1112-1121). IEEE.

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. Journal of business research, 69(2), 897-904.

European Commission. (2019). 2019 SBA Fact Sheet & Scoreboard, 1–52.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine*, *17*(3), 37-37.

Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). BDA capabilities and knowledge management: impact on firm performance. *Management Decision*.

Frąckiewicz-Wronka, A., & Szymaniec, K. (2012). Resource based view and resource dependence theory in decision making process of public organisation-research findings= Teoria zależności od zasobów oraz podejście zasobowe w procesie podejmowania decyzji w organizacji publicznej-wyniki badań empirycznych.

Google Trends. (2021). Retrieved 10 March 2021, from https://trends.google.com/trends/explore?date=all&geo=US&q=Big%20Data,Data%20Science,Data%20Analytics

Gilgun, J. F. (2019). Deductive qualitative analysis and grounded theory: Sensitizing concepts and hypothesis-testing. *The SAGE handbook of current developments in grounded theory*, 107-122.

Hazen, B. T., Skipper, J. B., Ezell, J. D., & Boone, C. A. (2016). Big data and predictive analytics for supply chain sustainability: A theory-driven research agenda. Computers & Industrial Engineering, 101, 592-598.

Hillman, A., Withers, M., & Collins, B. (2009). Resource Dependence Theory: A Review. Journal Of Management, 35(6), 1404-1427. doi: 10.1177/0149206309343469

Hou, T. (2021). IaaS vs PaaS vs SaaS: What You Need to Know + Examples (2018). Retrieved 15 May 2021, from https://www.bigcommerce.com/blog/saas-vs-paas-vs-iaas/#the-three-types-of-cloud-computing-service-models-explained

Iqbal, M., Kazmi, S. H. A., Manzoor, A., Soomrani, A. R., Butt, S. H., & Shaikh, K. A. (2018, March). A study of BD for business growth in SMEs: Opportunities & challenges. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 1-7). IEEE.

ISO 27001 Requirements. (2021). Retrieved 20 April 2021, from https://www.isms.online/iso-27001/requirements/

Kalan, R. S., & Ünalir, M. O. (2016, October). Leveraging BD technology for small and medium-sized enterprises (SMEs). In *2016 6th International Conference on Computer and Knowledge Engineering (ICCKE)* (pp. 1-6). IEEE.

Karim, S., Al-Tawara, A., Gide, E., & Sandu, R. (2017, May). Is BD too big for SMEs in Jordan?. In *2017 8th International Conference on Information Technology (ICIT)* (pp. 914-922). IEEE.

Koev, A. (2021). Transmetrics in logistics. Retrieved 19 April 2021, from https://thesiliconreview.com/magazine/profile/transmetrics-is-a-predictive-optimization-saas-for-the-transport-and-logistics-industry

Kuada, J. (2009). *Paradigms in International Business Research*. Working Paper Series.

Lavie, D. (2006). The competitive advantage of interconnected firms: An extension of the resource-based view. *Academy of management review*, *31*(3), 638-658.

Lebied, M. (2021). What Is Data Interpretation? Definition, Methods & Examples. Retrieved 11 May 2021, from https://www.datapine.com/blog/data-interpretation-methods-benefits-problems/#:~:text=Data%20interpretation%20refers%20to%20the,determines%20its%20signification%20and%20implications.

Lincoln, Y. S., & Guba, E. G. (1985). Establishing trustworthiness. *Naturalistic inquiry*, *289*(331), 289-327.

Lisy, K., & Porritt, K. (2016). Narrative synthesis: considerations and challenges. *International Journal of Evidence-Based Healthcare*, *14*(4).

Madasamy, M. (2021). Introduction to recommendation systems and How to design Recommendation system,that resembling the…. Retrieved 12 May 2021, from https://madasamy.medium.com/introduction-to-recommendation-systems-and-how-to-design-recommendation-system-that-resembling-the-9ac167e30e95

Maroufkhani, P., Ismail, W. K. W., & Ghobakhloo, M. (2020). BDA adoption model for small and medium enterprises. *Journal of Science and Technology Policy Management*.

Myers, M. D. (2009). Qualitative research in business & management. Sage, London

Mbassegue, P., Escandon-Quintanilla, M. L., & Gardoni, M. (2016, July). Knowledge management and BD: Opportunities and challenges for small and medium enterprises (SME). In *IFIP International Conference on Product Lifecycle Management* (pp. 22-31). Springer, Cham.

Nemati, A. R., Bhatti, A. M., Maqsal, M., Mansoor, I., & Naveed, F. (2010). Impact of resource based view and resource dependence theory on strategic decision making. International journal of business and management, 5(12), 110.

Noonpakdee, W., Phothichai, A., & Khunkornsiri, T. (2018, April). BD implementation for small and medium enterprises. In *2018 27th Wireless and Optical Communication Conference (WOCC)* (pp. 1-5). IEEE.

OECD. (2005). *OECD SME and entrepreneurship outlook 2005*.

OECD, S., & Outlook, E. (2019). Policy Highlights.

Open source. (2021). Retrieved 18 May 2021, from https://opensource.com/resources/what-open-source

Park, J., Fables, W., Parker, K., & Nitse, P. (2010). The Role of Culture in Business Intelligence. *International Journal Of Business Intelligence Research, 1*(3), 1-14. doi: 10.4018/jbir.2010070101

PayScale. (2021). Retrieved 13 May 2021, from https://www.payscale.com/research/DK/

Pfeffer, J., & Salancik, G. (2015). External Control of Organizations—Resource Dependence Perspective (pp. 373-388). Routledge.

Platform - Transmetrics. (2021). Retrieved 19 April 2021, from https://www.transmetrics.ai/platform/

Prahalad, C. K., & Hamel, G. (1997). The core competence of the corporation. In Strategische Unternehmungsplanung/Strategische Unternehmungsführung (pp. 969-987). Physica, Heidelberg.

Prasad, S., Zakaria, R., & Altay, N. (2018). Big data in humanitarian supply chain networks: A resource dependence perspective. Annals of Operations Research, 270(1), 383-413.

Predictive Empty Container Management. (2021). Retrieved 19 April 2021, from https://www.transmetrics.ai/case-study/assetmetrics-for-predictive-empty-container-management/#cs4

Predictive Planning - Case Study. (2021). Retrieved 19 April 2021, from https://www.transmetrics.ai/case-study/predictive-linehaul-planning-in-express-parcel-business/#cs3

Prescott, J. E., & Gibbons, P. T. (1993). The seven seas of global competitive intelligence. *Competitive Intelligence Review*, *4*(2-3), 4-11.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to BD and data-driven decision making. *BD*, *1*(1), 51-59.

Rajabion, L. (2018, December). Application and Adoption of BD Technologies in SMEs. In *2018 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1133-1135). IEEE.

Reitter, C. (2021). How to structure a powerful data team at smaller businesses — ReconInsight. Retrieved 11 May 2021, from https://www.reconinsight.com/blog/business-analytics-data-team-small-business

Ridzuan, F., & Zainon, W. M. N. W. (2019). A review on data cleansing methods for big data. *Procedia Computer Science*, *161*, 731-738.

Rothaermel, Frank T. (2013). Strategic Management: Concepts & Cases. New York: Irwin McGraw-Hill.

Runkler, T. A. (2016). Data Analytics. Models and Algorithms for Intelligent Data Analysis (Vol. I). *Munich, Germany: Springer. doi*, *10*, 978-3.

Russom, P. (2011). BDA. *TDWI best practices report, fourth quarter*, *19*(4), 1-34.

Ryan, G. W., & Bernard, H. R. (2000). Data management and analysis methods. Handbook of qualitative research, 2, 769-802.

Saunders, M., Lewis, P., Thornhill, A. (2009). Research methods for business students. Pearson Education Limited. 5th edition.

Shah, S., Soriano, C. B., & Coutroubis, A. D. (2017, December). Is BD for everyone? the challenges of BD adoption in SMEs. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)* (pp. 803-807). IEEE.

Sistla, D. (2021). 5 Big Data and Hadoop Use Cases in Retail Analytics. Retrieved 18 May 2021, from https://www.dezyre.com/article/5-big-data-and-hadoop-use-cases-in-retail-analytics/91

Spotify's algorithm. (2021). Retrieved 13 May 2021, from https://dittomusic.com/en/blog/how-does-spotifys-algorithm-work-streaming-hacks-for-musicians/

Stake, R. E. (1995). The Art of Case Study Research. Thousand Oaks, CA: Sage.

Statista Research department. (2021). Global BD market size 2011-2027 | Statista. Retrieved 31 March 2021, from https://www.statista.com/statistics/254266/global-big-data-market-forecast/

Swoboda, B., & Foscht, T. (2014). International Small and Medium-Sized Enterprises. *Trier, Alemania: Springer Gabler*.

Tian, R. G., & Tobar, B. G. (2004). Cross cultural strategies for competitive intelligence. *Competitive Intelligence Magazine*, *7*(5), 17-22.

Tozzi, C. (2021). Big Data Examples & Success Stories: Firefighting, Running and Beyond. Retrieved 12 May 2021, from https://www.precisely.com/blog/big-data/beyond-big-data-examples-success

Transmetrics. (2021). Retrieved 17 April 2021, from https://www.cloudwatchhub.eu/transmetrics-%E2%80%93-improving-cargo-capacity-utilisation-power-big-data-predictions

Transmetrics.ai. (2021). Retrieved 17 April 2021, from https://www.transmetrics.ai/

Transmetrics - A cloud solution. (2021). Retrieved 19 April 2021, from http://www.cloudwatchhub.eu/sites/default/files/Transmetrics.pdf

Transmetrics in Amsterdam. (2021). Retrieved 17 April 2021, from https://ajot.com/news/transmetrics-opens-new-commercial-headquarters-in-amsterdam

Transmetrics Project. (2021). Retrieved 31 May 2021, from https://transmetrics.eu/h2020-project/about-us/transmetrics-project/

Tudor, N. (2021). Blog | 7 real-world examples of how brands are using Big Data. Retrieved 25 May 2021, from https://www.bornfight.com/blog/7-real-world-examples-of-how-brands-are-using-big-data-analytics/

Ulrich, P., Becker, W., Fibitz, A., Reitelshöfer, E., & Schuhknecht, F. (2018). Data Analytics Systems and SME type–a Design Science Approach. *Procedia Computer Science*, *126*, 1162-1170.

Valcheva, S. (2021). Sustainable Competitive Advantage. Retrieved 13 May 2021, from http://www.intellspot.com/list-competitive-advantages/

Van Der Aalst, W. (2016). Data science in action. In *Process mining* (pp. 3-23). Springer, Berlin, Heidelberg.

VPN usage. (2021). Retrieved 20 April 2021, from https://www.namecheap.com/vpn/how-does-vpn-virtual-private-network-work/

Wang, S., & Wang, H. (2020). BD for small and medium-sized enterprises (SME): a knowledge management model. *Journal of Knowledge Management*.

Weiss, S.M. and N. Indurkhya (1998), Predictive Data Mining: A Practical Guide, Morgan Kaufmann Publishers, Inc

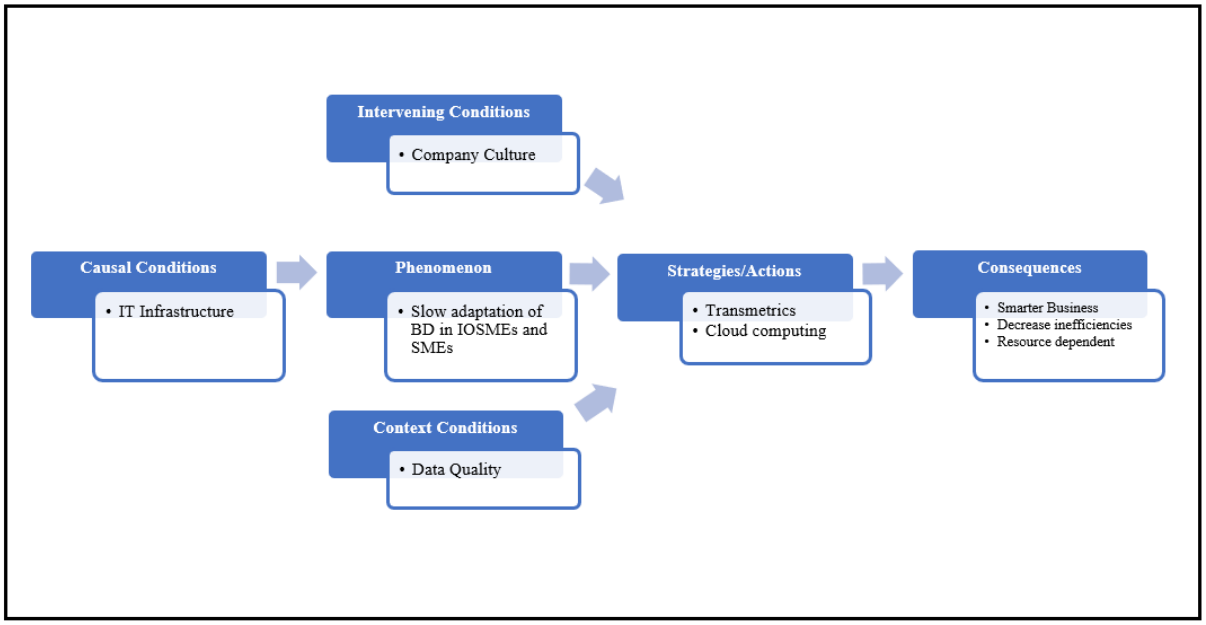
Wernerfelt, B. (1984). A resource-based view of the firm. Strategic management journal, 5(2), 171-180.

Yin, R. K. (2009). Case study research: Design and methods. Thousand Oaks, Calif: Sage Publications.

Appendix

Appendix 1: Adapted Axial framework for Transmetrics.

Own compilation, based on Corbin & Strauss (1990).



Appendix 2: Adapted Axial framework for AtScale

Own compilation, based on Corbin & Strauss (1990).